

# nature

THE INTERNATIONAL WEEKLY JOURNAL OF SCIENCE

SKYWATCHER

MIND MARVEL

INVISIBLE HAND

DOC QUANTUM

CAPTAIN MEDICA

CONTROL-X

BIOLOGENE

ENVIRONMENT

## SOMETHING IN THE AIR

Links between pollution and  
premature mortality

PAGES 330 & 367

ETHNOGRAPHY

## LIFE WITHOUT BORDERS

Mixed legacy of Victorian  
explorer Richard Burton

PAGE 319

MEDICINE

## SELF-DEFENCE COURSE

New technologies can make  
sense of human immunity

PAGE 409

## INTERDISCIPLINARITY

Why scientists must  
work together to save  
the world **PAGE 305**

NATURE.COM/NATURE

17 September 2015 £10

Vol. 525, No. 7569





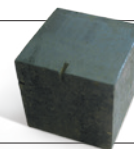
# THIS WEEK

## EDITORIALS

**OVERSIGHT** More institutions must check on animal experiments **p.290**

**WORLD VIEW** My unhappy time as an integrated social scientist **p.291**

**GONE FISSION** Nazi uranium never saw the inside of a reactor **p.292**



## Too close for comfort?

*Relationships between industry and researchers can be hard to define, but universities and other institutions must do more to scrutinize the work of their scientists for conflicts of interest.*

What sort of industry connections could buy influence over a scientist's research results? Research grants as small as US\$5,000? Money to support outreach that bolsters the industry's image? Equity in a spin-off company founded by the scientist? Defining what constitutes a conflict of interest — much less regulating it — continues to vex funding agencies, journals and institutions. Last month, for instance, *Nature* revealed that an activist organization had filed freedom-of-information requests to see the e-mails of researchers who work on genetically modified crops (see *Nature* 524, 145–146; 2015). Among other findings, their haul revealed that plant scientist Kevin Folta at the University of Florida in Gainesville had accepted a no-strings-attached \$25,000 grant from the agriculture giant Monsanto.

In his defence, Folta argued that the money supported only travel and outreach, not research, and he was therefore under no obligation to disclose it. This seems to be consistent with his institution's guidelines, and there is no evidence of any wrongdoing or that his research was compromised.

Solar physicist Willie Soon, a climate-change sceptic at the Harvard-Smithsonian Center for Astrophysics in Massachusetts, also seems to have been operating within institutional policy when advocacy groups revealed in February that he had accepted more than \$1 million from the energy industry, among other funders. (However, his failure to disclose those relationships might have violated the policies of some journals in which he published; see *Nature* <http://doi.org/2jx> (2015).)

In trying to navigate such complexities, the US National Institutes of Health (NIH) has been ahead of the curve — presumably because of long-standing concerns about physicians' industry relationships and the high stakes for protecting patients. Its parent agency, the Department of Health and Human Services (HHS), was the first to establish conflict-of-interest disclosure rules in 1995 and is still beyond many of its counterparts in maintaining unified regulations that include yearly reports to the government. By contrast, as one example, the US National Science Foundation's grants policy suggests that institutions look to scientific societies for ideas on how to manage a conflict of interest, and to report back to the foundation only if institutions cannot handle it themselves.

But even the HHS rules were not enough to guarantee full transparency. In 2009, a congressional report and subsequent media coverage found that some NIH-funded researchers had quietly accepted millions of dollars from industry. Again, the blame kept shifting: the universities said that the researchers had not reported the conflicts, the NIH received only bare-bones reports from institutions, and the researchers said that they did not know they were breaking any rules.

The HHS updated its policies in 2011, but pleased no one. The government underestimated the time and money that institutions would spend implementing new rules. And some aspects of the reforms have proved to be window dressing: a *Nature* investigation this week reveals that these reforms have uncovered few conflicts of interest that would have escaped the original regulations (see page 300).

The reforms may not be perfect, but they address real issues and others should take note. They make it clear that institutions are accountable, that they must educate their researchers on financial disclosure and that they should evaluate whether an industry relationship is problematic. The reforms also enlist a second pair of eyes by requiring institutions to report details of the conflict and its management to the NIH. Perhaps most importantly, the reforms remove the excuse of plausible deniability by clearly stating the kinds of financial relationship that could be considered conflicts.

***"The reforms may not be perfect, but others should take note."***

One thing has become clear: conflicts are slippery to define, so it is important for as many funders, institutions and journals to make as many demands as necessary. Had Kevin Folta been funded by the NIH, the HHS guidelines would have required him to report the Monsanto money. And if Willie Soon had had an NIH grant, his institution would have designed a 'management plan' that could have required his industry relationships to be stated in publications and lectures.

The HHS rules could backfire. Institutions do not want the publicity and work that accompany an identified conflict. Because they hold the power to decide whether a relationship presents a conflict, they could theoretically give their researchers a pass. *Nature's* investigation suggests that institutions use vastly different standards to evaluate such relationships, meaning that the rule is unevenly applied. And the current system makes it difficult for the public to access the conflict reports.

Still, the HHS should be commended for at least attempting to address the problem, even if it was forced into doing so. Other funders and institutions could do worse than to learn from its successes and mistakes if they define and strengthen their own policies. ■

## Mind meld

*Interdisciplinary science must break down barriers between fields to build common ground.*

In Castlegar, Canada, there is a golf shop that also offers vacuum-cleaner repairs, and in the Czech Republic town of Kostelec nad Orlicí, a business will sell you both wine and underwear. Such odd couplings are humorous because of their curiously limited scope.



**INTERDISCIPLINARITY**  
A *Nature* special issue  
[nature.com/inter](http://nature.com/inter)

There is nothing funny, after all, about a megastore that repairs equipment and sells golf



clubs, wine, underwear and everything else under the Sun.

The binary combinations also lead us to assume something about the shop's owners. Faced with a specific set of circumstances, these businesses redefine what we expect from a shop and offer something distinct.

There are greater problems in the world than what to do with your vacuum cleaner while you decide what make of balls to buy, but the principle is worth remembering as you browse this week's special issue of *Nature*, which we dedicate to interdisciplinary science.

Most scientists are aware of the term, and many will have used it. But how many are truly engaged in it? Done correctly, it is not mere multidisciplinary work — a collection of people tackling a problem using their specific skills — but a synthesis of different approaches into something unique. It is the wine and underwear shop, not the hypermarket.

The best interdisciplinary science comes from the realization that there are pressing questions or problems that cannot be adequately addressed by people from just one discipline. Witness the gathering of the scientific tribes — and the merging of approaches — for the Manhattan Project to work on the atomic bomb. More recently, *Nature* has reported on 'implementation science', which combines medical expertise with local knowledge on how best to carry out programmes to improve public health (see *Nature* 523, 516–518; 2015).

An interdisciplinary approach should drive people to ask questions and solve problems that have never come up before. But it can also address old problems, especially those that have proved unwilling to yield to conventional approaches.

Enough of the rhetoric, what about the reality? It is hard to deny that the scientific system — from funding streams and academic rewards to university departments and journals — does not encourage much

overlap between disparate subjects. It is easy to set up a 'Centre for Interdisciplinary Research', but who will be prepared to join it? If governments, funders and universities want to encourage more basic researchers to leave their trenches, then they need to make the no-man's-land of interdisciplinarity a more welcoming place to build a career. The obstacles are many, as we discuss in the pages that follow.

Some groups have found ways to overcome these obstacles, and some high-quality interdisciplinary work is under way. What are the key lessons from these successes?

**"True interdisciplinary science cannot be rushed."**

Interdisciplinary science takes longer than conventional projects, and that makes it more expensive. Funders must accept and embrace this and hold their nerve if the pay-off from individual projects takes longer than expected.

True interdisciplinary science cannot be rushed, not least because the best course of investigation is rarely clear at the outset. Research questions must be assessed and decided with input from all involved. An interdisciplinary project cannot exist as one main subject that sucks in the majority of the resources and leaves the partners as orbiting satellites.

Communication is crucial. The varying use of language across disciplines might seem a superficial problem, but it is one that must be solved, or misunderstandings will undermine the foundations of the project. There must also be no hierarchy, or perceived hierarchy. All involved must be confident that colleagues from other disciplines use equal academic rigour and scientific standing, even if the methods used in rival fields seem alien. It takes time to see the value in other approaches. It takes an open mind to appreciate an appliance-mending golf shop. ■

## Protection priority

*All involved in animal research must ensure that rules for ethical experiments are observed.*

**M**ore than a million people in Europe signed a petition earlier this year to halt research with animals. One reason why *Nature* and many scientists are able to defend these experiments is that all involved do everything they can to minimize pain and suffering. Animal experiments are approved only after thorough discussion and are carried out according to strict regulatory controls. Society sees the benefits of animal research, but it does not seek them at any cost.

When breaches of the strict rules that govern animal research occur, it is vital — to both supporters and opponents — that they are investigated thoroughly, and that lessons are learnt and shared. This week, *Nature* publishes a correction on its website that details such a breach of experimental protocol in a previously published paper (L. Raj *et al.* *Nature* <http://dx.doi.org/10.1038/nature15370>; 2015).

The relevant experiments grew tumours in mice as a way to test possible treatments. This type of study is common, as is the way they are approved and regulated. Researchers typically plan the experiments and then submit details to an institutional review board for approval. In making its decision, the board follows guidelines set out by a separate body charged with oversight of animal procedures — an institutional animal care and use committee. These guidelines are country-specific, and in the case of tumour experiments should include limits on the maximum tumour size allowed, and instructions to the researchers to monitor both tumour size and signs of distress.

In this case, prompted by a complaint from a reader and following consultation with the authors and the relevant bodies, *Nature* has established that the scientists did not carry out the required

monitoring properly. As a result, some of the tumours grew larger than permitted. These mice could therefore have experienced more pain and suffering than originally allowed for.

As well as writing to correct their paper to mark the breach of animal-welfare guidelines, the authors apologize for the breach. They are right to do so. Cases such as this could provoke a justifiable backlash against animal research. All involved — scientists, institutions, funders and journals — must do more to ensure that regulations are strictly observed.

*Nature's* policy is that the corresponding author on a paper that reports experiments with animals must confirm that the research was carried out in accordance with the relevant rules (see [go.nature.com/a9pjym](http://go.nature.com/a9pjym)). As a result of this case, we are increasing the amount of information we request from authors. In experiments in which tumours are grown, we now require authors to include the maximal tumour size permitted by the institutional animal-use committee, and to state that this was not exceeded. Authors must also provide the source data for any figures that analyse tumour growth.

*Nature* does not want to publish the results of experiments that have not been performed under ethical guidelines. As such, the authors in this case are correcting their paper to withdraw the portion of the data collected in experiments that the institutional committee concluded were in breach. The scientific conclusions of the paper remain valid and useful, and still stand.

Institutions should do more to make sure that the guidelines they set are respected. At the very least, on completion of each project — and before data are submitted — institutions should verify that approved protocols were followed. Funders and institutions must consider better training for young researchers doing work with animals. And the broader community should continue to scrutinize and improve how it carries out these types of experiment. Discussions are already

under way, for example, on whether the control arms of similar cancer studies truly need to let (untreated) tumours grow as large as they currently do. *Nature* is happy to join these discussions and to help to improve practice. ■

**➔ NATURE.COM**  
To comment online,  
click on Editorials at:  
[go.nature.com/xhunq](http://go.nature.com/xhunq)





## Integration of social science into research is crucial

*Social scientists must be allowed a full, collaborative role if researchers are to understand and engage with issues that concern the public, says Ana Viseu.*

Funders and institutions increasingly prioritize research that addresses the challenges and opportunities of an inherently interdisciplinary world. Policymakers and influential voices in science — including *Nature* — have also warned of a worrying disconnect between research and the needs and concerns of the public. One proposed solution is the integration of social scientists such as myself into publicly funded research initiatives. This is expected to contribute to the production of ‘better’ science.

Not in my experience. I spent three years as an in-house social scientist at the Cornell NanoScale Science and Technology Facility in Ithaca, New York, and the US National Nanotechnology Infrastructure Network, and it was a futile and frustrating time. I left a decade ago, but friends and colleagues who have since worked on similar projects tell me that the problem is widespread and that little has changed. Too many in the physical and life sciences dismiss social sciences as having a ‘service’ role, being allowed to observe what they do but not disturb it.

In its current model, integration is fuelled by the assumption that projects bring in the social sciences to carve a place for ‘society.’ This is expected to maximize the benefits of research while reducing negative impacts and public controversy. In other words, rather than being scientists in our own right, we are brought along as silent partners whose job it is to care for science. Rather than blurring boundaries and labour divisions, integration works to reify them. Thus, the questions that social scientists ask and the expertise we can contribute are muted or made invisible because we remain outside ‘proper’ science.

Integration is also deeply asymmetrical. The social sciences (often a single social scientist) are typically brought in after the project has taken shape. This asymmetry is present in every aspect of integration — from power to personnel numbers, funding, knowledge production and, ultimately, independence — but remains hidden in mundane interactions that dictate what counts as a valid social-science activity and who gets to define it.

This is not genuine integration. It pays lip service to the idea and is a waste of everyone’s time and the public money that supports it.

When I began my work alongside the nanotechnology scientists, I naively expected that my expertise as an ethnographer would be useful. I was prepared to study the culture of a laboratory and to probe its interaction with wider society. I thought that this would be helpful, given the frequent statements made by nanotechnology experts about how they wanted to engage and talk about the risks and benefits of their work.

Instead, the other scientists seemed to view my role as one of managing a narrow list of possible

risks and consequences, so that if a researcher followed my instructions and ticked boxes, then I would bless them as ‘social and ethical’ and they would be free to do their work with no concerns. I was routinely (wrongly) introduced as an ethicist and was expected to find minimal, non-disruptive ways of dealing with social and ethical issues. This was not a job that I could do nor wanted to do. Worse, my attempts to build bridges with my technical colleagues, for example by donning a cleanroom suit and learning how to use some of the equipment, were classified in lab annual reports as ‘outreach’. My perceived contribution was not one of expertise, but rather of a willingness to be educated in the proper way of thinking about nanotechnology.

Although my experience has left me sceptical of integration, I am not ready to dismiss the idea of fruitful collaboration between the natural and social sciences. Some fixes could be easily implemented: initiatives aiming for integration should have teams of social scientists, instead of one or two individuals, and these teams should be given the financial and operational autonomy to define and implement their activities.

When integration is planned, there should be a reassessment of what social scientists call the ‘positionality’ of the projects, which determines who pays for the research and thus who has the power to decide what is done, how it is done and what can be said about it.

For the social sciences to make meaningful contributions, funding structures must also be rethought. Ideally, we would see increases in stand-alone funding for social-science strands without requirements for integration or subordination to a topic. But this seems unlikely. Therefore,

we must push for project funding structures that — from the start — allocate and ring-fence money for the social-science component.

But this is not enough. For ‘integration’ to be productive, we must change its very meaning, from one of service to collaboration between equals. Doing so involves changes to scientific education and practice as well as continued reframing of our definitions of success. We must insist on the value of complexity, so that divergent thinking is not eclipsed in the effort to speak with one voice. We must make room for the disputes that are at the centre of knowledge production.

This is all the more important because, in a world of decreased funding for social sciences and humanities, speaking out of tune is both difficult and crucial. So we must begin to think of new means of partnership that will benefit us all. ■

Ana Viseu is associate professor at the Universidade Europeia in Lisbon, and a member of the Centro Interuniversitário de História das Ciências e Tecnologia, Universidade de Lisboa de Ciências, University of Lisbon. e-mail: [ana@anaviseu.org](mailto:ana@anaviseu.org)

FOR THE SOCIAL  
SCIENCES TO MAKE  
MEANINGFUL  
CONTRIBUTIONS,  
FUNDING  
STRUCTURES MUST BE  
RETHOUGHT.

➔ [NATURE.COM](http://NATURE.COM)  
Discuss this article  
online at:  
[go.nature.com/3u8qe9](http://go.nature.com/3u8qe9)



# RESEARCH HIGHLIGHTS

Selections from the  
scientific literature

## COMMUNICATION

### Climate sceptics use strong words

Climate scientists use more cautious language in scientific reports than do climate-change sceptics, even though the sceptics often accuse the scientists of being alarmist.

Srdan Medimorec and Gordon Pennycook at the University of Waterloo in Canada used software to analyse the style of language in a report by the Intergovernmental Panel on Climate Change (IPCC) in 2013 and in a response written by a sceptic group, the Nongovernmental International Panel on Climate Change (NIPCC). The researchers did not assess the scientific accuracy of the reports but found that the NIPCC report used emotional language and the IPCC report contained more neutral and formal phrasing.

The authors hypothesize that the IPCC uses such language because of scrutiny from the media and sceptics. *Clim. Change* <http://doi.org/7mb> (2015)

## NUCLEAR PHYSICS

### Forensics reveals uranium's past

Uranium from German experiments during the Second World War was not used in a nuclear reactor for any appreciable amount of time.



Maria Wallenius at the European Commission Joint Research Centre's Institute for Transuranium Elements in Karlsruhe, Germany, and her colleagues did a forensic analysis of uranium samples (pictured) used in 1940s experiments in Germany. They looked for trace elements and isotopes of uranium and plutonium that are created when neutrons released during nuclear fission smash into other atoms.

They traced the origin of the uranium to a mine in the Czech Republic, and

found that isotope ratios matched those found in natural uranium ore. The samples were never used in experiments that reached the critical mass necessary for sustained nuclear fission. *Angew. Chem. Int. Ed.* <http://doi.org/f3f7js> (2015)

## CANCER

### A trap for roving cancer cells

Implanting a polymer scaffold in mice that have tumours captures spreading cancer cells,

enabling their early detection.

Lonnie Shea at the University of Michigan in Ann Arbor and his colleagues placed human breast-cancer cells in mice and implanted the scaffolds in their abdomens a week later. Two weeks after cell transplantation, the researchers detected cancer cells in the scaffolds but not in the lungs or liver, where breast cancer often spreads. After 28 days, mice with scaffolds had fewer tumours in their lungs than did animals without scaffolds. And using an imaging technique, the team measured changes in the tissue



## ANIMAL BEHAVIOUR

### Whales that click create cliques

Sperm whales form clans by learning vocal calls from others that sing like them. This kind of 'cultural transmission' has been seen as a mainly human trait.

Sperm-whale clans use distinct dialects of clicks to communicate. To learn how their complex societies form, Maurício Cantor at Dalhousie University in Halifax, Canada, and his colleagues used 18 years of data on the acoustic calls of sperm whales (*Physeter macrocephalus*; pictured) from around the Galapagos Islands to build several possible

models of whale populations. In their simulations, the clans that have been observed in nature did not form when the vocal calls were genetically inherited or learned from other sperm whales in general. But clans did form when the animals adopted the most common calls produced by certain individuals — mainly those with similar communication patterns.

This further suggests that humans are not the only mammals that segregate according to similarities in learned behaviour.

*Nature Commun.* 6, 8091 (2015)

FLIP NICKLIN/MINDEN PICTURES/FLPA

EUROPEAN COMMISSION



properties within the scaffold that indicated the presence of cancer cells.

An inflammatory response to the scaffold attracted the cancer cells. This approach could eventually be used in humans to detect the early spread of cancer, the authors say.

*Nature Commun.* 6, 8094 (2015)

## PLANETARY SCIENCE

## A faster spin for Mercury

Mercury rotates nine seconds faster than scientists had thought, probably because of gravitational effects from Jupiter.

A team led by Alexander Stark of the German Aerospace Center in Berlin studied three years of data from NASA's MESSENGER spacecraft, which orbited the planet between 2011 and 2015 and measured Mercury's rotations more precisely than ever before.

The data also confirm that the planet has a molten outer core, causing this part to rotate at a different speed from the solid inner layers.

*Geophys. Res. Lett.* <http://doi.org/7mc> (2015)

## CANCER

## Muscle wasting blocked in mice

Giving tumour-bearing mice specific proteins prevents a muscle-wasting syndrome that commonly affects people with cancer.

Many patients with cancer die from severe muscle loss (cachexia), which has no treatment. To find a way to halt the condition, Amelia Johnston and Nicholas Hoogenraad at La Trobe University in Melbourne, Australia, and their colleagues injected mice with mouse cancer cells that had been engineered to express a human gene encoding the protein Fn14, which drives cancer growth. The animals lost muscle and fat, but giving the mice an antibody against

Fn14 stopped cachexia. Moreover, in a mouse model of cachexia, the animals lived longer and maintained body weight when treated with an anti-Fn14 antibody, compared with untreated mice.

Targeting Fn14 proteins that are generated by tumours could be a treatment strategy for this condition, the authors say. *Cell* 162, 1365–1378 (2015)

## ASTRONOMY

## The farthest galaxy so far

Astronomers have observed the most distant galaxy yet by detecting photons emitted from its clouds of hydrogen when the 13.8-billion-year-old Universe was less than 600 million years old.

Such photons rarely make it to telescopes on Earth, but Adi Zitrin at the California Institute of Technology in Pasadena and his colleagues were able to detect them using a telescope at the W. M. Keck Observatory in Mauna Kea, Hawaii. They found that the wavelength of arriving photons had been stretched en route, indicating that the galaxy, named EGSY8p7, is more than 13.2 billion light years (4 billion parsecs) away.

Seeing hydrogen emission from such a distant galaxy may challenge current understanding of the evolution of the Universe, the authors say.

*Astrophys. J. Lett.* 810, L12 (2015)

## ECOLOGY

## Marauding ants bring disease

One of the most widespread invasive ant species not only displaces native ants, but also carries viruses.

Phil Lester at Victoria University of Wellington and his colleagues searched

## SOCIAL SELECTION

Popular topics on social media

### Science failings shared on Twitter

Researchers' best success stories end up in journals, but many of their less-successful ones found their way on to Twitter this week with the hashtag #FailingInSTEM. Tales of low points and often-humorous mishaps reassured others that failures can be overcome on the way to scientific success. "The #FailingInSTEM tweets are so important! It's so comforting to know that other scientists make mistakes," tweeted Aimee Eckert, a PhD student in cell biology at the University of Sussex in Brighton, UK. Nicole Cabrera Salazar, an astronomy PhD student at Georgia State University in Atlanta, started the #FailingInSTEM Twitter discussion after a friend of hers suffered a scientific setback: "We need to let our young ppl know that regular, fallible people do science. We make mistakes everyday. It's part of the job #FailingInSTEM." She suspected that other young researchers could use a reminder that science is not all about

➔ **NATURE.COM**  
For more on popular papers:  
[go.nature.com/mzblhl](http://go.nature.com/mzblhl)

successful experiments and flashy publications. "People don't talk about all of the times that they broke something in a lab or got heckled during a presentation," she says.

for viral sequences in RNA extracted from Argentine ants (*Linepithema humile*; pictured) in New Zealand. They found a virus that they named Linepithema humile virus 1, which could explain periodic crashes in Argentine ant populations. They also found that the ants carried deformed wing virus, which can be fatal to honeybees.

The team suggests that bees could become infected when the ants forage or raid bee nests. *Biol. Lett.* 11, 20150610 (2015)

## CONDENSED-MATTER PHYSICS

## Weyl particles discovered

Three separate teams have found analogues of Weyl fermions: massless elementary particles that were first predicted in 1929 but

have never been observed.

Physicists searching for these fermions look for their unusual properties in the collective behaviour of other particles. Hong Ding and Tian Qian at the Chinese Academy of Sciences in Beijing and their colleagues saw these 'quasiparticles' by probing a sample of tantalum arsenide with a beam of X-rays. In July, a separate group of researchers led by Zahid Hasan at Princeton University in New Jersey announced that they had seen the particles in the same material. Ling Lu at the Massachusetts Institute of Technology in Cambridge and his colleagues reported seeing signs of the particles in the behaviour of light passing through a crystal.

Such experimental systems could allow researchers to probe the exotic properties associated with Weyl particles. *Phys. Rev. X* 5, 031013 (2015); *Science* 349, 613–617; 622–624 (2015)

➔ **NATURE.COM**  
For the latest research published by Nature visit:  
[www.nature.com/latestresearch](http://www.nature.com/latestresearch)





# SEVEN DAYS

The news in brief

## POLICY

### Clone-products ban

The European Parliament has voted in favour of a sweeping ban in the European Union on the use of food and feed products — domestic or imported — from cloned animals and their descendants. The rules tighten a draft law proposed in 2013 by the European Commission to prohibit the sale of food products derived from cloned animals. Parliamentarians said that the proposed amendments, voted in on 8 September, reflect widespread consumer concerns over animal welfare and food safety. But Vytenis Andriukaitis, the European commissioner for health and food safety, called the amendments “disproportionate” and warned that they might prove “legally impossible”.

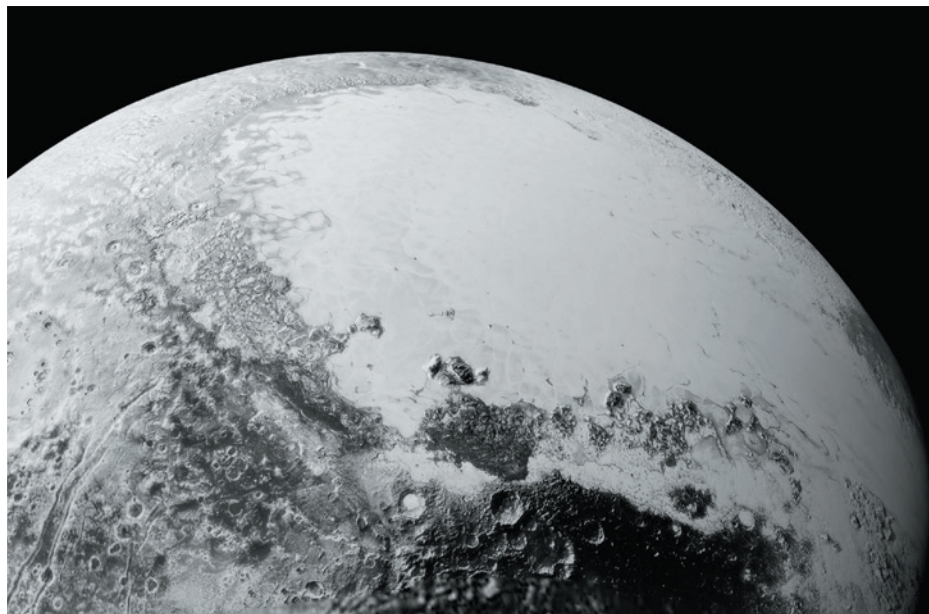
### Energy review

The US Department of Energy released its second Quadrennial Technology Review on 10 September, which looks at current energy technologies and identifies opportunities for research and development. The report suggests that the US energy system is becoming increasingly diverse with the rise of renewable power, as well as being more interconnected through

## NUMBER CRUNCH

# 879

**Total number of days Russian cosmonaut Gennady Padalka has spent in space, the most by any individual. Padalka returned from his latest stay on the International Space Station on 11 September.**



NASA/JHU/APL/SRI

## Pluto's 'heart' snapped in high resolution

The left lobe of Pluto's bright heart-shaped feature, Tombaugh Regio, is clearly seen in the upper right of this image released on 10 September. The view spans 1,800 kilometres and is generated from high-resolution images

gathered during the 14 July fly-by of Pluto by NASA's New Horizons spacecraft. The boundary between the bright, icy plains (called Sputnik Planum) and dark, cratered terrains (called Cthulhu Regio) is particularly striking.

Internet and communications technologies; both trends open the door to cleaner energy and fewer emissions. Although the United States has made progress on energy efficiency, the report notes that substantial opportunities remain for reducing energy consumption and costs.

### Sonar muffled

The US Navy has agreed to limit its sonar and explosives activities in areas that might harm dolphins, whales and other marine mammals. The agreement between the navy and environmental groups, including Earthjustice and the Natural Resources Defense Council, was ordered by a federal judge on 14 September. The navy will not be able to

use sonar, which can disrupt communication between marine mammals, in areas off the Southern California coast. Areas around Hawaii's islands are protected from sonar and explosives training operations.

### Energy ambitions

The California legislative assembly passed laws on 11 September that will increase requirements for renewable-energy production in the state. The bill, sought by Governor Jerry Brown, raises the current renewable-energy quota of 33% by 2020 to a more-rigorous 50% by 2030. It also sets a goal of doubling energy efficiency in the electricity and natural-gas sectors by 2030. An earlier draft of the bill would have required the state to halve its oil use

over the same period, but the provision was dropped because of feasibility and cost concerns.

## RESEARCH

### Land undervalued

Global land degradation costs between US\$6.3 trillion and \$10.6 trillion per year, according to the Economics of Land Degradation (ELD) Initiative in Bonn, Germany. In a report published on 15 September, the ELD said that unsustainable land management ruins productivity and removes ecosystem services that have no market value, including nutrient recycling and disease regulation. The loss figure includes the costs of replacing these services, for



example by buying fertilizers or vaccines. People without funds to replace services such as clean water are particularly vulnerable to land degradation.

## Permafrost tracked

The first international database of standardized permafrost data was launched this week by the Global Terrestrial Network for Permafrost (GTN-P), an international consortium that aims to establish an early-warning system for permafrost thawing, for use by scientists and policymakers. The European Union-funded database gathers frozen-soil temperatures and annual thaw depths. Permafrost has a key role in climate-change modelling, because when it thaws it can release the greenhouse gases carbon dioxide and methane.

### EVENTS

## Turnbull coup

Malcolm Turnbull has ousted his fellow Liberal party member Tony Abbott as Australian prime minister, after forcing a party ballot on 14 September. Turnbull (pictured) won the leadership vote by 54 votes to 44. On his first day, Turnbull said that he would keep current climate-change policies for now. In 2014, Abbott repealed



Australia's carbon tax and scrapped an emissions-trading scheme, disappointing Australian climate scientists. In 2009, Turnbull was overthrown as party leader by Abbott, in part because of Turnbull's support for an emissions-trading scheme.

## Pesticide repealed

A US appeals court has rescinded the approval by the Environmental Protection Agency (EPA) of an insecticide named sulfoxaflor. The approval was challenged by a group of bee-keeping organizations, which cited evidence that sulfoxaflor — a neonicotinoid compound with an unusual mechanism — is highly toxic to bees. The court ruled on 10 September that the EPA's decision was based on 'flawed and limited data'. Dow AgroSciences, which won approval for sulfoxaflor in 2013, is considering challenging the ruling.

### BUSINESS

## Phage trial starts

A European company has started the first randomized clinical trial using bacterium-killing viruses called phages. On 9 September, Pherecydes Pharma, based in Romainville, France, announced that it had begun enrolling people with burns who are susceptible to infection by the bacteria *Escherichia coli* and *Pseudomonas aeruginosa*. The company will test two cocktails of phages that attack these species in trials involving 220 people in Switzerland, France and Belgium. Phage therapy has been used for decades in Eastern Europe, but has not yet been tested in a large, controlled trial.

### FUNDING

## Boost for Africa

Researchers in African nations will share more than £46 million (US\$70 million) in a programme to build science capacity. The first seven 'Developing Excellence in Leadership, Training and Science' (DELTA) awards were announced on 10 September. They range from mental-health research in Zimbabwe to science-leadership training in Kenya. Funded by the London-based biomedical charity the Wellcome Trust, together with the UK Department for

## COMING UP

### 17–18 SEPTEMBER

Scientists from across academia and industry gather in Nairobi to discuss biotechnology and biomedical science in Africa.

<http://aibbc.org/>

### 17–21 SEPTEMBER

The 30th International Papillomavirus Conference convenes in Lisbon, offering workshops in clinical and public health.

<http://www.hpv2015.org/>

### 21–25 SEPTEMBER

Sandbjerg, Denmark, hosts the 3rd International Workshop on Microbial Life under Extreme Energy Limitation.

<http://microenergy2015.org/>

International Development and the Bill & Melinda Gates Foundation, the DELTAs programme will from 2016 be managed by the newly formed Alliance for Accelerating Excellence in Science in Africa (see *Nature* 520, 142–143; 2015).

## New sponsor sought

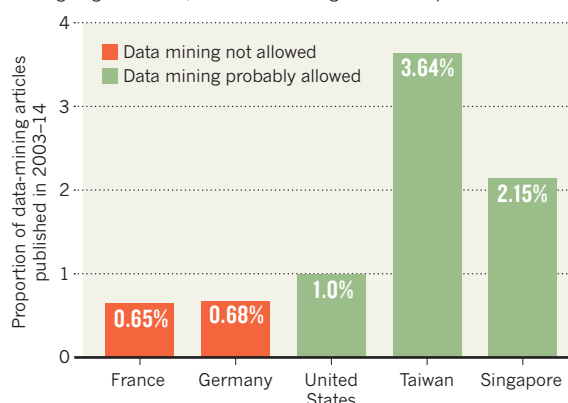
Intel will drop its support for the oldest and most prestigious science contest for US high-school students, Science Talent Search, after the 2017 event. Eight Nobel prizewinners, five winners of the National Medal of Science and many other scientists are alumni of the 73-year-old programme, in which students are judged on original research. The Society for Science & the Public in Washington DC, which runs the competition, announced on 9 September that it was looking for a new sponsor after Intel pulled its US\$1.6-million annual backing, begun in 1998.

## TREND WATCH

Researchers in countries that often give academic exemptions from copyright laws, such as the United States, publish more data-mining studies than in places such as Germany and France, where academics must first gain consent. The work, presented at the 2 September European Policy for Intellectual Property meeting in Glasgow, UK, analysed 18,441 articles on data mining, a method that allows investigators to trawl large data sets for discoveries (C. Handkey *et al. Soc. Sci. Res. Netw.* <http://doi.org/7px; 2015>).

## COPYRIGHT IMPACTS ON DATA MINING

In countries where copyright restrictions are relaxed for researchers mining large data sets, more data-mining articles are published.



# NEWS IN FOCUS

**POLICY** NIH conflict-of-interest rules get poor reviews **p.300**

**PHYSICS** LIGO resumes hunt for gravitational waves **p.301**

**ASTRONOMY** Security cameras capture 86 new meteor showers **p.302**

**COLLABORATIONS** The most interdisciplinary fields and countries are revealed **p.306**



WITS UNIVERSITY



Lee Berger (front) recruited a team of wiry excavators to retrieve more than 1,500 fossils from the Dinaledi chamber in South Africa.

## PALAEOANTHROPOLOGY

# Crowdsourcing digs up an early human species

*Palaeoanthropologist asks excavators and anatomists to study Africa's richest fossil trove.*

BY EWEN CALLAWAY

“Dear colleagues — I need the help of the whole community,” palaeo-anthropologist Lee Berger posted on social media on 6 October 2013.

Berger, based at the University of Witwatersrand in Johannesburg, South Africa, had just learnt of a small underground chamber loaded with early human fossils. He was

looking for experienced excavators to collect the delicate remains before they deteriorated further. “The catch is this,” Berger went on. “The person must be skinny and preferably small. They must not be claustrophobic, they must be fit, they should have some caving experience.”

Less than two years after he posted this missive, Berger and his team have pieced together more than 1,500 ancient human bones

and teeth from the Rising Star cave system — the biggest cache of such material ever found in Africa. The remains belong to at least 15 individuals of a previously undescribed species that the team has dubbed *Homo naledi*, and they may mark the oldest-known deliberate burial in human history, Berger and his colleagues report in *eLife* (L. R. Berger *et al.* *eLife* **4**, 09560; 2015 and P. H. G. M. Dirks *et al.* *eLife* **4**, 09561; 2015). For Berger, the research marks a milestone in ▶



► a campaign to transform palaeoanthropology into an open and inclusive field, in which rare fossils are rapidly shared with the scientific world instead of being squirrelled away as an elite few scrutinize them for years.

“There’s lots of fossils out there no one has ever seen, except for a few select people. Palaeoanthropology is really rotten that way,” says Tracy Kivell, a palaeoanthropologist at the University of Kent in Canterbury, UK, who analysed hand bones from Rising Star and is a co-author of the paper that describes *H. naledi*. “Lee is changing that and setting a new standard for what we should expect.”

Palaeoanthropologist Denné Reed of the University of Texas at Austin sees Berger’s openness as part of a generational shift in the field. “We’re more interested in openly sharing data,” he says. “The advantages in collaboration far outweigh any of the risks.”

A few weeks before Berger advertised for help, cavers who work with him had discovered the Dinaledi Chamber in the Rising Star cave system, about 50 kilometres northwest of Johannesburg. Berger hoped to remove the remains as soon as possible, but he needed help. The narrow chamber is about 30 metres below ground, and the only access is through a slit in the rock some 20 centimetres wide (see ‘Tough commute’). “I’m not physiologically appropriate to ever get in the system,” he says, referring to his large build. A month after his social-media post, Berger had six scientists at work in the cave.

### TREASURE TROVE

As he and other colleagues watched a video feed in a nearby tent, the excavators pulled up skulls, femurs, teeth and hundreds of other specimens. “By the end of this expedition, we had recovered more individual remains than had been discovered in South Africa in the last 90 years,” says Berger.

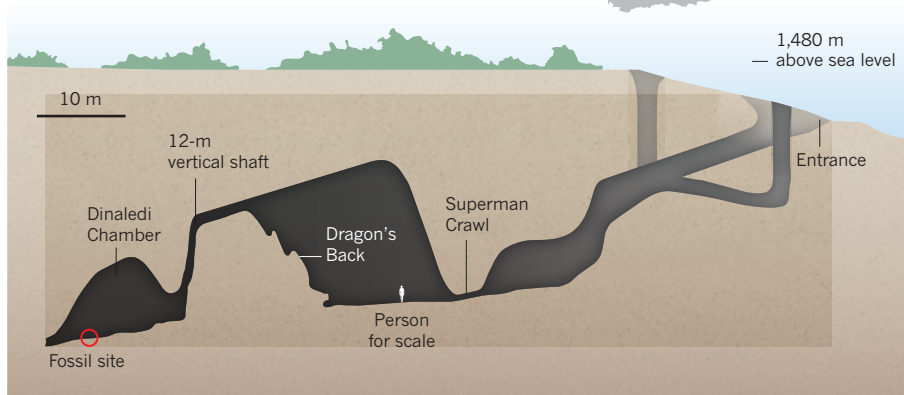
The research team with which Berger usually works to analyse early human remains was busy making sense of other fossils, discovered in 2008 at the nearby Malapa site. So Berger put out another social-media call, this time recruiting more than 30 early-career scientists to attend a month-long workshop to analyse and describe the fossils.

John Hawks, a palaeoanthropologist at the University of Wisconsin–Madison who helped to coordinate the Rising Star dig and workshop, says that the team took flak for its unorthodox approach. “There’s a lot of the field that really believed we’re just a couple of cowboys who don’t know how things should be done,” he says.

The team intends to publish at least a dozen papers from the workshop in the coming months; the two published today are the first. They describe the discovery site and the anatomy of *H. naledi*, the skull of which encased a small, fist-sized brain much like those of other early members of the genus *Homo* and of the

### TOUGH COMMUTE

To access the fossil-rich Dinaledi Chamber, six slender scientists had to penetrate deep into the Rising Star cave system, squeezing themselves through a passage dubbed Superman Crawl because cavers have to extend one arm overhead — like the flying superhero — to get through.



more ancient australopiths. In other ways, its body is more like those of modern humans, with the lower limbs and feet of a biped and hands that could have gripped tools with precision. The researchers estimate that *H. naledi* would have stood just under 1.5 metres tall and weighed between 40 and 55 kilograms.

“It is a very strange combination of features, some that we’ve never seen before and some that we would have never expected to find together,” says Hawks.

### FAMILY RESEMBLANCE

The researchers are unclear about how *H. naledi* is related to other early human species that lived in Africa, such as *Homo erectus* and *Homo habilis*. They hope to date calcite



*Homo naledi*’s skull looks like an australopithecine’s.

deposits in the cave to establish the age of the remains, which could be more than 1 million years old.

The chamber contains no evidence that early humans lived there, and no bones from species other than *H. naledi*, so Berger believes that it might be a deliberate burial, and possibly the oldest known. Currently, the oldest site that seems to represent an early human burial is Sima de los Huesos in the Atapuerca Mountains of Spain, which dates to 430,000 years ago.

Fred Spoor, a palaeontologist at University College London, agrees that the bones represent a previously unknown *Homo* species, and says that Berger’s team makes a good case, after considering other alternatives, that the remains were deposited intentionally. He is eager to see what other experts make of it.

However, Jeffrey Schwartz, an evolutionary biologist at the University of Pittsburgh in Pennsylvania, thinks that the material is too varied to represent a single species. “I could show those images to my students and they would say that they’re not the same,” he says. One of the skulls looks more like it comes from an australopithecine, he adds, as do certain features of the femurs.

Schwartz and others will soon get the chance to judge the Rising Star remains for themselves. Berger’s team has uploaded data including 3D scans of the remains to the MorphoSource repository, and welcomes other researchers to study the material at first hand. Berger did the same with remains of a species called *Australopithecus sediba* that were discovered at the Malapa site.

Schwartz says that he has had trouble accessing some researchers’ hominin remains even after they had been described in a journal. But when he asked Berger’s team if he could purchase *A. sediba* casts several years ago, he got them for free. “How good can you be?” says Schwartz. “It’s been refreshing and delightful that Lee Berger has always made his specimens accessible.” ■

SOURCE: P. H. G. M. DIRKS ET AL. *ELIFE* 4, 09561 (2015)

WITS UNIVERSITY

## GLOBAL HEALTH

# Africa braced for snakebite crisis

Health specialists warn that stocks of antivenom will run out in 2016.

BY QUIRIN SCHIERMEIER

Rural Africa is facing a resurgence of a persistent plague that rarely makes headlines: snakebite.

By June next year, stockpiles of the antivenom that is most effective against Africa's vipers, mambas and cobras are expected to run out because the only company that makes the medicine has stopped production. With no adequate replacement in sight, the death toll from bites is set to rise, specialists warned at a tropical-medicine congress last week in Basel, Switzerland.

"We're dealing with a neglected health crisis that is turning into a tragedy for Africa," says Gabriel Alcoba, a medical adviser with the international humanitarian group Médecins Sans Frontières (MSF; also known as Doctors Without Borders).

Poisonous snakes might seem an archaic menace in such a rapidly urbanizing world. Yet by cautious estimates, snakebites kill more than 100,000 people worldwide every year (see 'Death toll') — more, on average, than lose their lives in natural disasters. And survivors often experience permanent physical and mental disabilities.

In 2010, the French drug firm Sanofi Pasteur in Lyon ceased production of Fav-Afrique, an antibody serum that reduces the quantity of venom circulating in the blood of a snakebite victim. Made from the purified plasma of horses previously injected with small quantities of



The deadly carpet viper (*Echis ocellatus*).

snake venom, the serum neutralizes the venom of many of Africa's most dangerous snakes.

The antidote has saved many people from bites by deadly species such as the carpet viper (*Echis ocellatus*), common in West Africa, and the black mamba (*Dendroaspis polylepis*), found across the sub-Saharan region. But the high costs — US\$250–500 per person — and a supply shortage mean that only about 10% of snakebite victims in Africa get treatment, and the company says that producing the antidote is no longer profitable. Cheaper products by competitors have forced Sanofi Pasteur out of the African market, says Alain Bernal, a company

spokesman. Sanofi Pasteur is working to enable the transfer of know-how to companies willing to take over production of Fav-Afrique, he says.

Pharmaceutical companies in South Africa, India, Mexico and Costa Rica are among those marketing cheaper products — some of which work well against snakes in their host nations. But their safety and effectiveness against the large variety of species in Africa have not yet been established in clinical trials. To speed up the process, MSF is offering two of its hospitals in the Central African Republic (CAR) and South Sudan as study sites. But it will take at least two years to validate the products in development, and none is as broadly efficient as Fav-Afrique, Alcoba says.

## NEGLECTED THREAT

Although just now becoming critical, Africa's snakebite problem has been smouldering for years, says tropical-medicine specialist David Warrell of the University of Oxford, UK, who consults for the World Health Organization (WHO). Snakebite fatalities have been rising over the past decade in the CAR, Ghana and Chad — in part owing to a failure to train enough medical staff, ignorance from health ministries and "unscrupulous marketing" of inappropriate antivenoms, he says. "War-torn countries have many other problems. But the millions of children, poor farmers and nomadic people at risk of snakebites just don't have the ear of politicians in capital cities."

And according to Warrell, the WHO has done little to help. To improve the safety and efficacy of antibodies, the agency has released guidelines for producing antivenoms. But it has no formal programme for improving treatment by training medical workers, advising ministries or educating communities, as it does for 17 other neglected tropical diseases, including dengue and sleeping sickness. And yet, says Warrell, snakebites cause more deaths than do all 17 diseases put together.

Warrell says that, while waiting for clinical trials to bring replacements for Fav-Afrique to the market, the keys to reducing the risk of snakebite are education and preventive measures — such as wearing proper shoes, using a light when walking home from the fields and sleeping above ground level, beneath a mosquito net.

Thankfully, says Alcoba, the global-health community is starting to grasp the urgency of the situation. "People used to laugh when we talked about snakebites," he says. "They don't laugh any more." ■

## DEATH TOLL

### Fuzzy estimates of snakebite fatalities

It is uncertain how many people are bitten or die from snakebites in sub-Saharan Africa. But according to Médecins Sans Frontières (MSF; also known as Doctors Without Borders), whose health-care workers treat snakebites through field programmes in the Central African Republic and South Sudan, an estimated 30,000 people die each year and at least 8,000 more undergo amputations.

But snakebite mortality could be much higher than anecdotal reports suggest. For some countries, including the Democratic Republic of Congo — home to an enormous number of venomous snakes — there are no reliable data, says tropical-medicine specialist David Warrell at the University of Oxford, UK.

Under-reporting is not limited to Africa. The authors of a nationally representative snakebite-mortality survey, published in 2011, deduced that, despite the availability of antidotes, around 46,000 people in India die of snakebites every year (B. Mohapatra *et al.* *PLoS Negl. Trop. Dis.* **5**, e1018; 2011). India's Central Bureau of Health Intelligence reported merely 1,219 and 985 fatal bites for 2009 and 2010, respectively. One reason for the discrepancy, says Warrell, who co-authored the study, is that many victims of snakebites die before they reach a hospital, or waste precious time with traditional healers before seeking more-conventional medical help. **Q.S.**



# NIH disclosure rules falter

Regulations that require researchers to disclose conflicts of interest yield questionable data and cost universities millions.

BY SARA REARDON

When a US Senate investigation in 2008 revealed that psychiatrist Charles Nemeroff of Emory University in Atlanta, Georgia, had not disclosed at least US\$1.2 million in income from drug companies, Senator Charles Grassley decided to do something about it. The Iowa Republican led a charge to push the National Institutes of Health (NIH), which funded Nemeroff's research, to change how it evaluates researchers who accept money from industry.

The resulting reforms, which took effect in 2012, require scientists to report industry connections in greater detail than before, and charge institutions with determining which ties are problematic. But three years later, it is not clear what the costly, cumbersome rules have accomplished. A *Nature* analysis suggests that institutions have vastly different standards for what constitutes a conflict — and that they classify relatively few relationships between researchers and industry as troublesome.

"There's a lot more financial conflict of interest in my view than the NIH is getting from the reports of universities," says Sheldon Krinsky, who studies conflict-of-interest issues at Tufts University in Medford, Massachusetts. "We're just seeing the tip of the iceberg."

The reforms, enacted by the NIH's parent agency, the Department of Health and Human Services (HHS), do seem to have increased the number of financial relationships that researchers report to their universities — by 45% overall, according to data from 56 universities in a survey released in April by the Association of American Medical Colleges (AAMC) in Washington DC (see [go.nature.com/hc5r2b](http://go.nature.com/hc5r2b)). But the number of conflicts that institutions reported to the NIH has increased only slightly, according to NIH data obtained by *Nature* through a freedom-of-information request (see 'Under the microscope').

The agency's original conflict-of-interest regulations, implemented in 1995, required institutions to report when an HHS-funded researcher received more than \$10,000 from an outside source. The revised rule lowered that threshold to \$5,000 and directed researchers to disclose a wider variety of potential conflicts, such as sponsored travel and relationships with non-profit organizations.

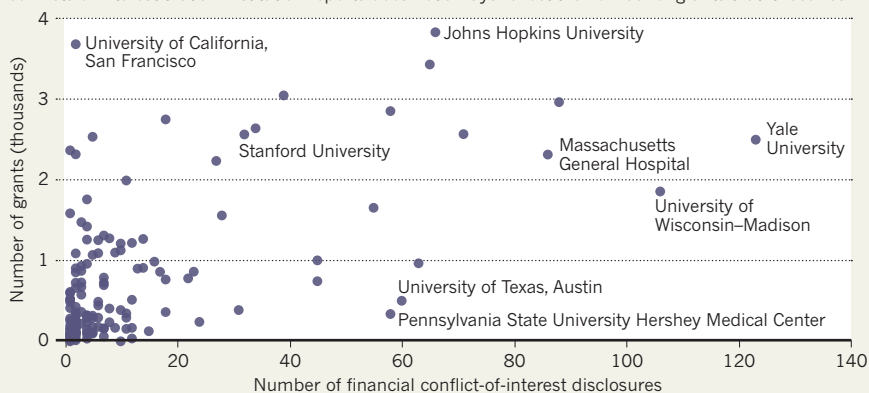
Institutions, which receive conflict-of-interest

## UNDER THE MICROSCOPE

Through a freedom-of-information request, *Nature* obtained conflict-of-interest reports submitted to the US National Institutes of Health (NIH). For more on our methodology, see [go.nature.com/11pjj6](http://go.nature.com/11pjj6)

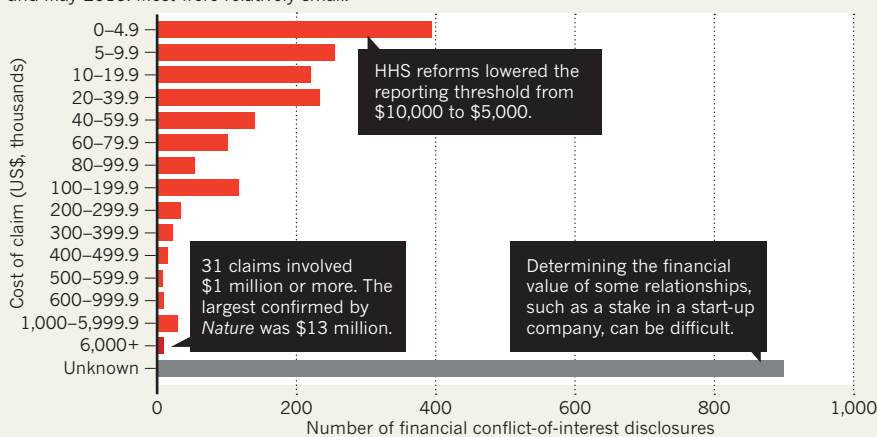
### OUTLOOK HAZY

Data from the NIH, which cover the period from August 2012 to May 2015, suggest that the number of conflicts of interest that an institution reports does not always reflect the number of grants that it receives.



### SMALL CLAIMS

Institutions reported 2,523 financial conflicts of interest between January 2013 and May 2015. Most were relatively small.



reports from their researchers annually, must then convene an internal panel to determine whether a particular relationship could affect a researcher's work. If so, the panel designs a 'management plan' that may require the researcher to disclose the conflict in publications or, in some cases involving human subjects, to stand down as the study's primary investigator. Institutions then send these plans to the NIH.

Universities have spent millions of dollars and hired extra staff to comply with these reforms, and most administrators are furious about the burden. "We already had an annual disclosure process for all the faculty," says

Andrew Rudczynski, associate vice-president for research administration at Yale University in New Haven, Connecticut. "I can't see a single benefit to it."

Yale spent \$500,000 to implement the revised NIH rules. In the year after they took effect, the number of disclosures by the university's researchers doubled — but Yale identified just one new conflict, Rudczynski adds. Other universities report similar experiences.

And whereas the HHS had estimated that the roughly 2,000 institutions that it funds would spend \$23.2 million a year to comply with the regulations, the AAMC survey suggests that the true cost has been much higher. Just

SOURCE: NIH

71 institutions spent a total of \$23 million in the year after the reforms took effect, although their costs going forward may be lower.

Paul Thacker, who led the 2008 Senate investigation as a member of Grassley's staff, admits that it is difficult to know how well the reforms are working. That is largely because the potential benefits of greater disclosure of financial ties, such as peer reviewers giving closer scrutiny to studies by researchers with conflicts, are tough to measure.

Still, Thacker says, there is a clear need for closer scrutiny. This is backed up by evidence showing that studies funded by private sources, such as drug firms, more often produce results that benefit the funder than do publicly funded studies (A. Lundh *et al.* *Cochrane Database Syst. Rev.* **12**, MR000033; 2012). And Thacker has little sympathy for universities' complaints. "It just shows that they still don't get what the problem is," he says. "They're in this place today because they've failed to create confidence for the public in the past."

Others worry that the HHS policy is still not strict enough. Krimsky says that the current rules may give institutions too much power to assess conflicts, without accounting for ways that universities themselves can be compromised by ties to government or industry. This could be one reason why the HHS reforms did not significantly increase the number of reported conflicts, Krimsky adds.

Those pushing for greater transparency are also frustrated that the NIH does not require institutions to publish information about researchers' conflicts and management

plans online. Instead, members of the public must ask a university for information on a researcher's conflicts; the institution has five days to disclose dollar amounts and sources. Nonetheless, the NIH Office of Extramural Research says that about 50% of institutions that submit conflict-of-interest reports have voluntarily created online databases, although these vary in usability and completeness.

Requesting such information from universities directly also produces mixed results. *Nature* contacted 20 public and private institutions that had reported individual researchers with conflicts of interest involving more than \$1 million, seeking details

**"We are still at the mercy of what's disclosed to us."**

on these relationships. The majority of these institutions responded immediately, but some took as long as two weeks to respond, directed *Nature's* reporter to the media office, or instructed her to submit a freedom-of-information request. Most declined to share information about conflicts that occurred before the current calendar year, which is not required by the HHS.

Nor does the department require the release of management plans, which troubles Tobin Smith, vice-president for policy at the Association of American Universities in Washington DC. "If you disclose that there is a conflict but don't disclose how the university is managing it — which is not part of the regulations — the public doesn't understand the relationship," he says.

The NIH also struggles to defend its own

regulations. "One could debate whether or not we needed to promulgate a new rule," says Sally Rockey, director of the NIH Office of Extramural Research. "At the time, there was a lot of scrutiny in the press and Congress got involved." She concedes that the reforms were mostly in response to this outside pressure. (Grassley declined to comment on the regulations.)

And it is unclear whether the revised regulations would have identified Nemeroff, who did not tell Emory about his industry relationships. "Science and research are built on trust, and we are still at the mercy of what's disclosed to us," says Eric Mah, senior director of research compliance at the University of California, San Francisco.

The NIH plans to review the conflict-of-interest reforms later this year, to develop best practices for compliance. The agency will examine data on the type and number of reported conflicts, as well as institutions' experiences of complying with the requirements. But Rockey says that the HHS is unlikely to make significant changes to the rules, given that they took four years to develop.

In the meantime, research institutions are caught in a bind. The 1980 law that allows US universities to patent inventions encourages relationships with industry, and tight federal research budgets are driving more scientists to seek support from private funders. "There are no easy answers," Thacker says. "Universities are being pushed into greater reliance on industry funding and until that reverses, these problems just become more and more complicated." ■

## PHYSICS

# Hunt for cosmic waves to resume

*Upgraded LIGO detectors will improve chances of finding ripples in space-time.*

BY DAVIDE CASTELVECCHI

**A**lmost 100 years after Einstein presented the general theory of relativity in a Berlin lecture theatre, the quest to spot the gravitational waves he predicted may be entering its final stages.

This week, the world's largest gravitational-wave facility is expected to start collecting data again after a 5-year US\$200-million overhaul. The Laser Interferometer Gravitational-Wave Observatory (LIGO) searched fruitlessly for these cosmic ripples for almost a decade in the 2000s. But the odds that its improved version — known as Advanced LIGO — will detect any waves in the next three months may be as high as one in three, according to some of the physicists involved in the experiments.

Initial tests have shown that the observatory's twin detectors, in Washington state and Louisiana, are performing as expected, says Gabriela González, spokesperson for the 900-strong LIGO Scientific Collaboration. And that is no mean feat for an instrument that has cost \$620 million so far. "It's the first time that anything in this field is on budget and on schedule," says Karsten Danzmann, director of the Max Planck Institute for Gravitational Physics, in Hannover, Germany, who is not part of the LIGO management team.

According to general relativity, gravitation originates from the interplay between massive objects and the malleable fabric of space-time. Einstein predicted that accelerating masses such as colliding neutron stars or black holes would disturb that fabric and produce gravitational

ripples that propagate through the Universe.

Each of LIGO's detectors is designed to measure the deformation of space-time by comparing changes in the paths of laser beams that race down its two perpendicular 4-kilometre-long arms, bounce between mirrors and interfere with each other back at their source. When a gravitational wave passes through, it slightly alters the lengths of the arms, and the observatory can spot such changes with a sensitivity of one part in  $10^{22}$ . That is comparable to a hair's-width change in the distance from the Sun to Alpha Centauri, its nearest star, says Laura Cadonati, a physicist at the Georgia Institute of Technology in Atlanta who will be coordinating the experiment's data analysis.

A crucial part of the improvement is better damping of the vibrations caused by ►



► less-than-heavenly sources. The problem was especially acute at the site in Livingston, Louisiana, which is in the middle of a timber plantation. Any felling of trees would disturb the detector, so it could keep its laser beams 'in lock' — vibrating at precise frequencies — only at night or on weekends. A passing train would knock the site out for an hour, says physicist Brian O'Reilly, who will coordinate the follow-up of detections at the Livingston site. But now, he says, the detector should be able to take data over several days at a time without interruption.

Advanced LIGO is already three times more sensitive than its predecessor, but in three months' time it will shut down for more improvements that will make it ten times more sensitive. When it reopens around 9 months later, it should be able to spot cosmic ripples from cataclysmic events — such as the collisions of black holes — up to 120 megaparsecs (326 million light years) away on a regular basis and sample a volume of space 1,000 times greater than the original observatory.

Next year, LIGO will be joined by a slightly smaller €200-million (US\$226-million) Franco-Italian detector near Pisa, Italy, called Advanced Virgo, which is undergoing its own upgrade. The LIGO and Virgo teams will pool their data to check each other's detections. They expect to see waves from mergers of binary neutron stars — events that should generate strong, predictable signals — but do not know precisely how many to anticipate. "It could be, depending on the models, ten binary neutron star detections a year or so," González says. "But it could be 10 times higher or 100 times lower."

"The first detections will be quite dramatic for us," says Rainer Weiss, a theoretical physicist at the Massachusetts Institute of Technology in Cambridge who was one of LIGO's founders. "The first thing we will need to sort out is whether we truly believe what we are seeing."

Having detectors on different continents is crucial for providing a rough estimate of the origin of the waves, says Fulvio Ricci, a physicist at the Sapienza University of Rome and the spokesperson for Virgo. Once they know that, astronomers will be able to look for other signs of that event using electromagnetic radiation, such as X-rays or visible light.

Einstein published his first papers on gravitational waves in 1916. Detecting these ripples a century later, Weiss says, would be of "enormous symbolic importance". ■



A meteor (upper left) streaks through the Orion constellation during the Perseid shower.

#### ASTRONOMY

# Dates added to meteor calendar

*Skywatching cameras spot 86 previously unknown events.*

BY ALEXANDRA WITZE

The list of meteor showers that occur every year has just grown longer. Eighty-six previously unknown showers have now joined the regular spectaculars, which include the Perseids, Leonids and Geminids. Astronomers spotted the shooting-star shows using a network of video cameras designed to watch for

burglars, but repurposed to spy cosmic debris burning up in Earth's atmosphere.

The newfound showers are faint but important: each is fuelled by Earth's passage through a trail of particles left behind by a comet or asteroid, so mapping them reveals previously unknown sources of dust.

"The cool thing is, we are not just doing surveillance of meteors in the night sky," says Peter Jenniskens, an astronomer at the

BABAK TAHERI/NATIONAL GEOGRAPHIC CREATIVE



**MORE  
ONLINE**

#### INTERVIEW



Why marine life needs protection from noise pollution  
[go.nature.com/qdldtz](http://go.nature.com/qdldtz)

#### MORE NEWS

- California snowpack lowest in past 500 years [go.nature.com/c2gul3](http://go.nature.com/c2gul3)
- Scientists trial humane shark deterrents [go.nature.com/uqlhl1](http://go.nature.com/uqlhl1)
- Southern Ocean sucks up more carbon dioxide than was thought [go.nature.com/wfb4vh](http://go.nature.com/wfb4vh)

#### NATURE PODCAST



Camouflaging drug delivery; science meets theatre; and the health impacts of air pollution [nature.com/nature/podcast](http://nature.com/nature/podcast)

SETI Institute in Mountain View, California. “Now we also have a three-dimensional picture of how dust is distributed in the Solar System.”

Most of the particles are the size of a sand grain, but a few are large enough to survive the searing heat of their passage through the atmosphere — and possibly do damage on Earth’s surface. Jenniskens and his colleagues describe the discoveries in four papers accepted for publication in *Icarus*.

Astronomers have been documenting meteors for centuries, first by eye and more recently with radar and video-tracking systems. Meteors sprinkle Earth steadily throughout the year, but during a shower a significant number seem to originate from the same point in the sky. Skywatchers around the world have reported more than 750 possible meteor showers to the International Astronomical Union (IAU) — but only a small fraction of those have been confirmed as bona fide events.

### SKY SURVEILLANCE

Jenniskens’ team set up cameras at three locations in northern California to confirm or rule out these rumoured showers. The Cameras for Allsky Meteor Surveillance (CAMS) project points 60 security cameras in different directions to capture as many shooting stars as possible. Each has a relatively narrow field of

view, but together they cover a broad dome of sky centred directly overhead and extending down to 30° above the horizon.

“CAMS is about getting massive data sets on meteors, so you can see through all the scatter to get at those new showers,” says Phil Bland, a planetary scientist at Curtin University in Perth, Australia. He helps to run a tracking network in the Australian

outback that looks for extremely bright meteors in an effort to recover meteorites on the ground.

Since it began in 2010, CAMS has measured more than 250,000 meteors.

Of those, about three-quarters were random singletons and one-quarter came in showers. CAMS has confirmed 81 showers that were on the IAU’s questionable list, and discovered 86 new ones.

Among these is one that lights up Southern Hemisphere skies in early December, and seems to radiate from the constellation Vela. It is surprisingly strong for a shower that had not been noticed before, says Jenniskens. During the March 2013 peak of a newly confirmed shower, skywatchers saw the bright flash of a rock-sized object hitting the Moon.

**“The more we sample the sky, the more detailed our picture becomes.”**

The CAMS team has been expanding its search by setting up smaller camera networks in the Netherlands and New Zealand. “The more we sample the sky,” says Jenniskens, “the more detailed our picture becomes of what is coming in.” ■

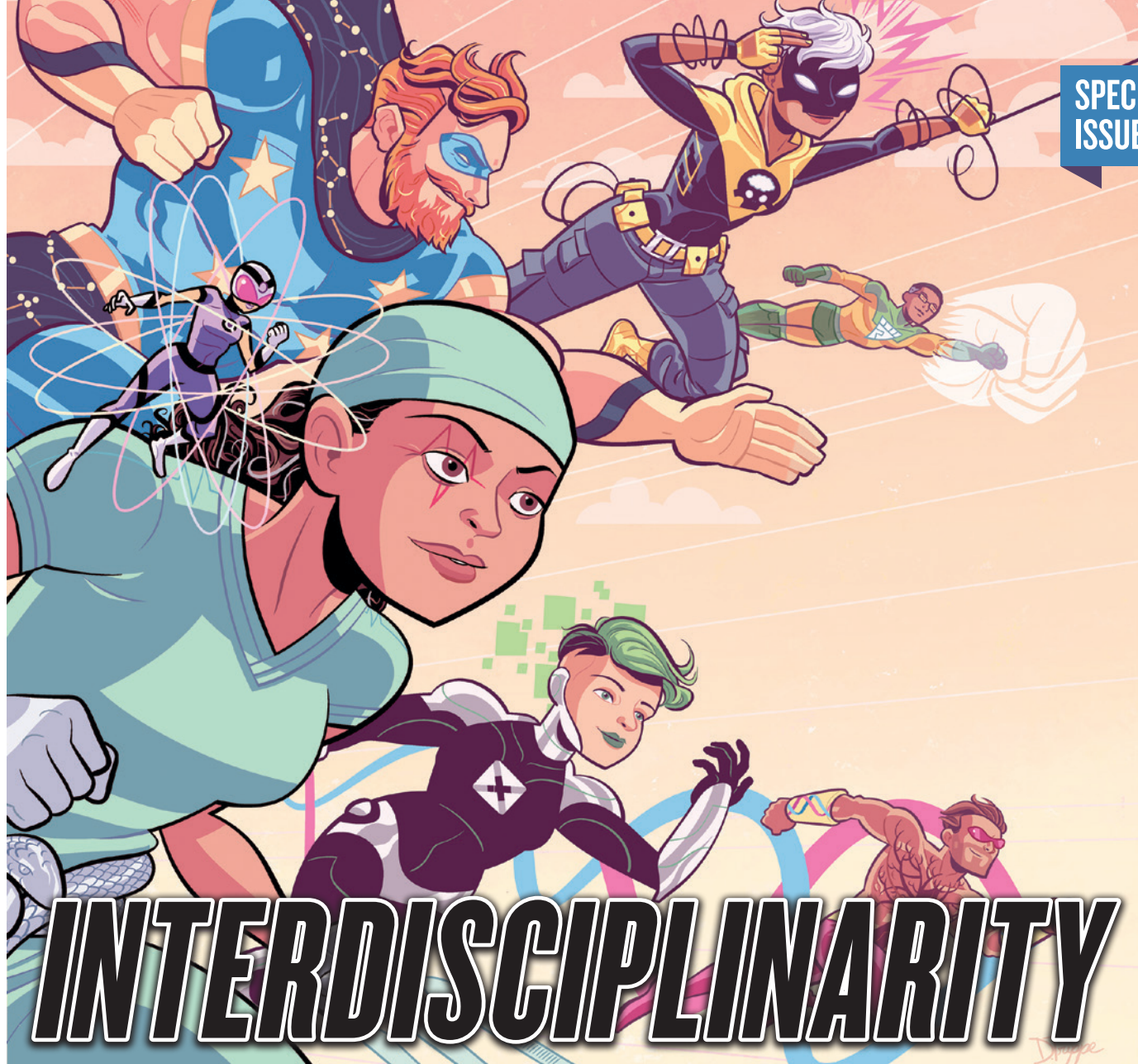
### CORRECTIONS

The News story ‘Encryption faces quantum foe’ (*Nature* **525**, 167–168; 2015) incorrectly named the location for the cryptography workshop that began on 6 September. The workshop was held at the Schloss Dagstuhl–Leibniz Centre for Informatics in Wadern, not the Leibniz Center for Informatics in Oktavie-Allee.

The News Feature ‘Fishing for the first Americans’ (*Nature* **525**, 176–178; 2015) incorrectly credited the photo taken at Cooper’s Ferry. Credit should have gone to Hayden Wilcox, not Joanne McSparran.

The News story ‘Health study set to decide data policy’ (*Nature* **525**, 16–17; 2015) incorrectly stated that an NIH working group planned to create a blanket data-sharing policy for the Precision Medicine Initiative. It is in fact developing a policy that can accommodate participants’ varying interest in seeing their own genetic information.





## Scientists must work together to save the world. A special issue asks how they can scale disciplinary walls.

**T**o solve the grand challenges facing society — energy, water, climate, food, health — scientists and social scientists must work together. But research that transcends conventional academic boundaries is harder to fund, do, review and publish — and those who attempt it struggle for recognition and advancement (see World View, page 291). This special issue examines what governments, funders, journals, universities and academics must do to make interdisciplinary work a joy rather than a curse.

A News Feature on page 308 asks where the modern trend for interdisciplinary research came from — and finds answers in the proliferation of disciplines in the twentieth century, followed by increasingly urgent calls to bridge them. An analysis of publishing data explores which fields and countries are embracing interdisciplinary research the most, and what impact

such research has (page 306). On page 313, Rick Rylance, head of Research Councils UK and himself a researcher with one foot in literature and one in neuroscience, explains why interdisciplinarity will be the focus of a 2015–16 report from the Global Research Council. Around the world, government funding agencies want to know what it is, whether they should invest in it, whether they are doing so effectively and, if not, what must change.

How can scientists successfully pursue research outside their comfort zone? Some answers come from Rebekah Brown, director of Monash University's Water for Liveability centre in Melbourne, Australia, and her colleagues. They set out five principles for successful interdisciplinary working that they have distilled from years of encouraging researchers of many stripes to seek sustainability solutions (page 315). Similar ideas help scientists, curators

and humanities scholars to work together on a collection that includes clay tablets, papyri, manuscripts and e-mail archives at the John Rylands Research Institute in Manchester, UK, reveals its director, Peter Pormann, on page 318.

Finally, on page 319, Clare Pettitt reassesses the multidisciplinary legacy of Richard Francis Burton — Victorian explorer, ethnographer, linguist and enthusiastic amateur natural scientist who got some things very wrong, but contributed vastly to knowledge of other cultures and continents. Today's would-be interdisciplinary scientists can draw many lessons from those of the past — and can take our polymathy quiz online at [nature.com/inter](http://nature.com/inter). ■



**INTERDISCIPLINARITY**  
A Nature special issue  
[nature.com/inter](http://nature.com/inter)

# INTERDISCIPLINARY RESEARCH BY THE NUMBERS

An analysis reveals the extent and impact of research that bridges disciplines.

BY RICHARD VAN NOORDEN

Interdisciplinary work is considered crucial by scientists, policymakers and funders — but how widespread is it really, and what impact does it have? Scholars say that the concept is complex to define and measure, but efforts to map papers by the disciplines of the journals they appear in and by their citation patterns are — tentatively — revealing the growth and influence of interdisciplinary research.

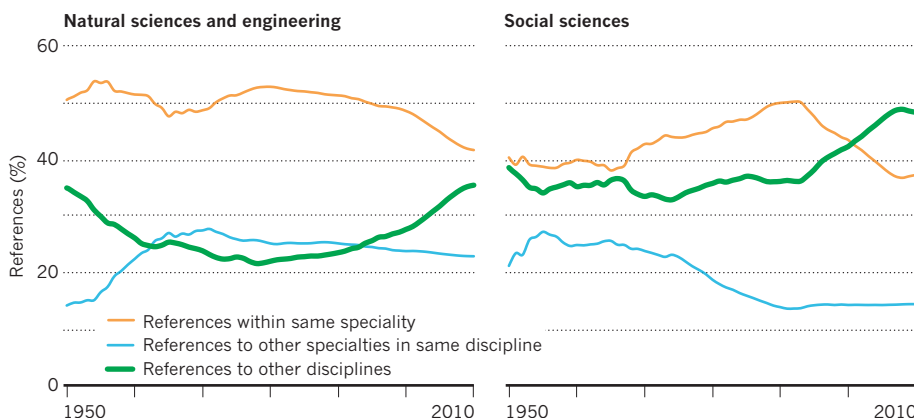


**INTERDISCIPLINARITY**  
A *Nature* special issue  
[nature.com/inter](http://nature.com/inter)

## 1 Interdisciplinary research is on the rise

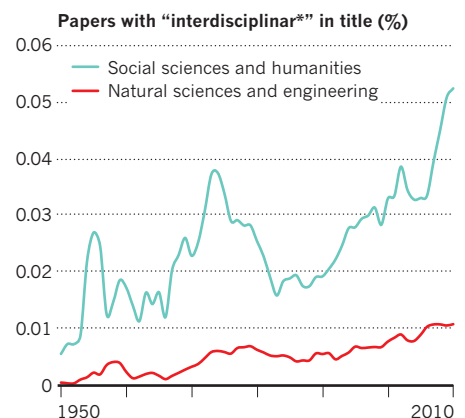
### REFERENCES

Since the mid-1980s, research papers have increasingly cited work outside their own disciplines. The analysis shown here used journal names to assign more than 35 million papers in the Web of Science to 14 major conventional disciplines (such as biology or physics) and 143 specialities. The fraction of paper references that point to work in other disciplines is increasing in both the natural and the social sciences. The fraction that points to another speciality in the same discipline (for example, a genetics paper pointing to zoology) shows a slight decline.



### RHETORIC

Discourse about interdisciplinary research is increasing. The fraction of papers that mention interdisciplinarity in their title has fluctuated, perhaps reflecting the priorities of funders, but the twenty-first century saw that proportion reach an all-time high.

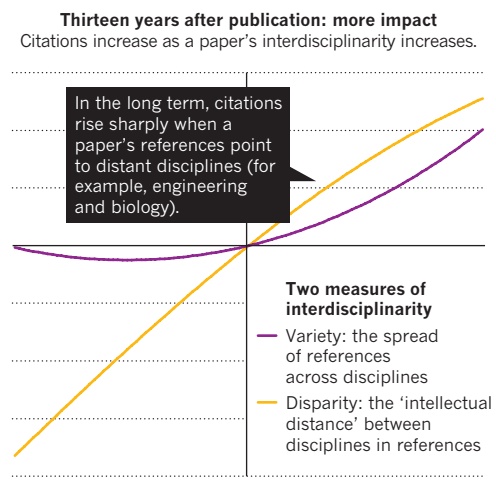
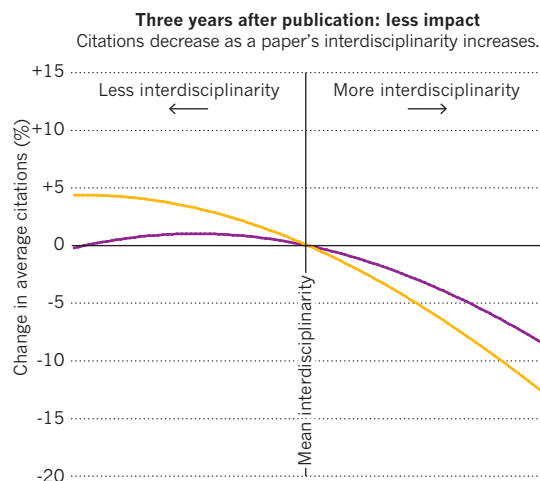


SOURCE: V. LARIVIERE & Y. GINGRAS IN *BEYOND BIBLIOMETRICS* (EDS B. CRONIN & C. R. SUGIMOTO) 187–200 (MIT PRESS, 2014)

## 2 Interdisciplinary research takes time to have an impact

### IMPACT

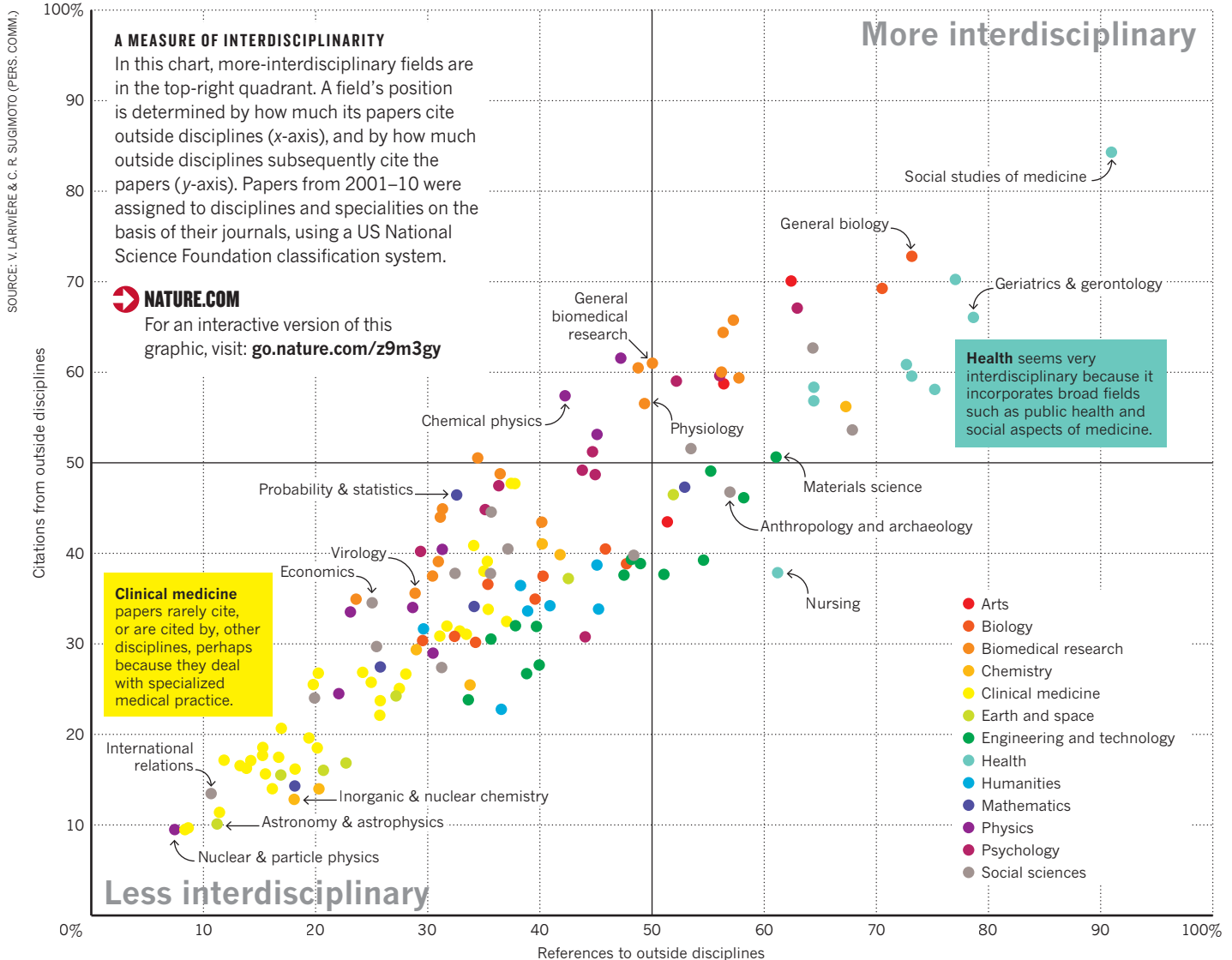
Whether interdisciplinary research gains more citations than disciplinary research is contentious. Over three years, papers with diverse references tend to pick up fewer citations than the norm, but over 13 years they gain more. Some studies suggest that a little interdisciplinarity is better than a lot: papers that combine very disparate fields tend to earn fewer citations. But interdisciplinary work can have broad societal and economic impacts that are not captured by citations.



SOURCE: J. WANG ET AL. *PLOS ONE* 10, E0127298 (2015)



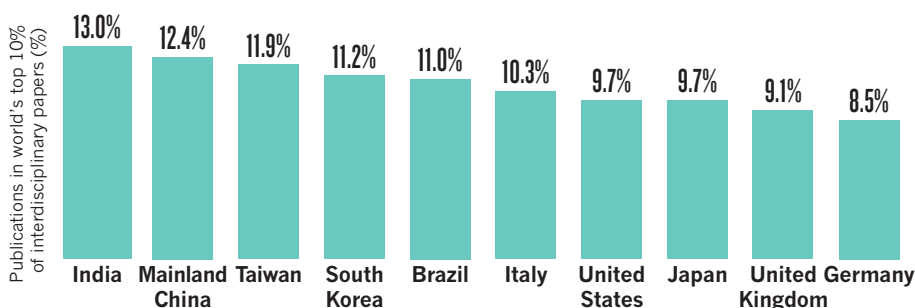
### 3 Some fields are more interdisciplinary than others ...



### 4 ... and so are some countries

#### MOST INTERDISCIPLINARY COUNTRIES

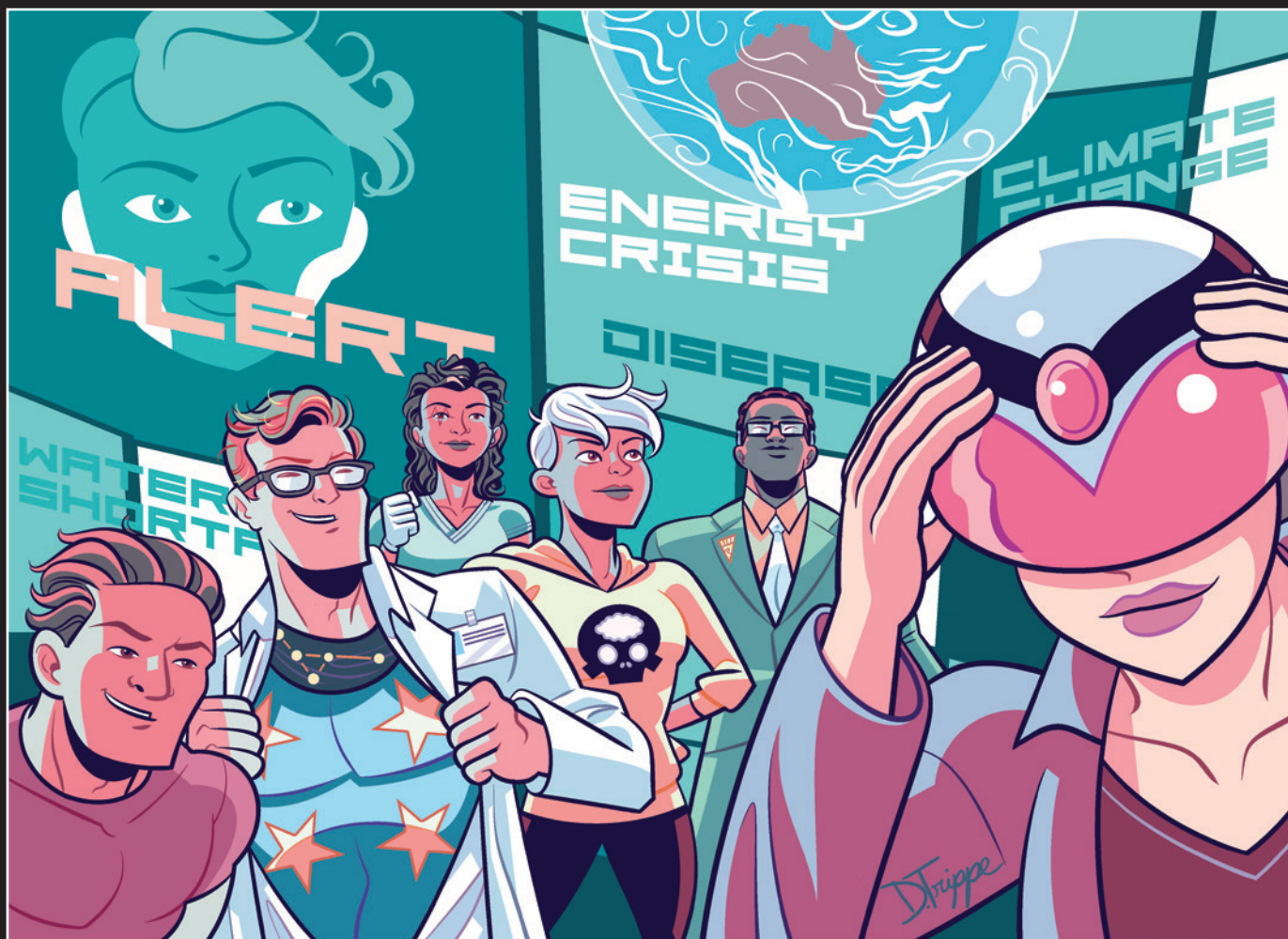
A 2015 study by researchers with the publisher Elsevier defined interdisciplinary papers as those that reference journals that are rarely cited together. The report looked only at countries that routinely publish more than 30,000 papers per year to find the 'most interdisciplinary' countries for 2013.



A separate analysis counted the proportion of a paper's references that are in other disciplines. After totting up all the papers for each country, and normalizing the results (so that average interdisciplinarity = 1), similar nations emerge on top for 2013.

1. Mainland China (1.09<sup>1</sup>)
2. India (1.07)
3. Taiwan (1.06)
4. Brazil (1.04)
5. Australia (1.02) and South Korea (1.02)

<sup>1</sup>9% higher than world's average interdisciplinarity



# TEAM SCIENCE

Interdisciplinarity has become all the rage as scientists tackle society's biggest problems. But there is still strong resistance to crossing borders.

BY HEIDI LEDFORD

Asking for US\$40 million is never easy, but Theodore Brown knew his pitch would be a particularly tough sell. As vice-chancellor for research at the University of Illinois at Urbana-Champaign in the early 1980s, Brown had been tasked with soliciting a major donation from wealthy chemist and entrepreneur Arnold Beckman, a graduate of the university. Beckman was hesitant, believing that the university should receive most of its support from the state. So Brown decided to devise a project like nothing he had ever seen before.

In 1983, he and his colleagues put together a proposal for an institute that had little chance of being funded through normal channels. It would defy the powerful disciplinary cartography that defines many modern universities, bringing together members of different departments and inducing them to work together on common projects. Brown argued that it would allow faculty members to tackle bigger scientific and societal questions than they normally could.

"The problems challenging us today, the ones really worth working on, are complex, require sophisticated equipment and intellectual tools, and just don't yield to a narrow approach," he says. "The traditional structure of university departments and colleges was not conducive to cooperative, interdisciplinary work."

It was an early example of the push for interdisciplinary research that is now sweeping universities around the globe. Although Brown was not completely alone — the interdisciplinary Santa Fe Institute in New Mexico was founded around the same time — he was advocating crossing boundaries before it

ILLUSTRATION BY DEAN TRIPPE



became fashionable. And his proposal met strong resistance. Department heads fretted that faculty members — and their grants — would be snatched away. Some colleagues scorned Brown's idea of creating open office spaces to foster interactions between graduate students: surely the din would make it impossible to get serious work done. And then there was the stigma. "Interdisciplinary research is for people who aren't good enough to make it in their own field," an illustrious physicist chided.

But Beckman liked the idea and committed the full \$40-million asking price — at that time, the largest-ever private donation to a US public university. A few hectic years later, the 29,000-square-metre Beckman Institute for Advanced Science and Technology was born.

The institute struggled to recruit a qualified director willing to take a chance on the new model, so Brown took the helm. Soon, large grants from organizations such as the Department of Defense and the National Science Foundation poured in, hushing many critics. By the time Brown left the institute in 1993, other leading universities were sending delegations there to learn from the model. Researchers from Beckman — which now has more than 200 affiliated faculty members — have achieved attention-grabbing results, including helping to create one of the first graphical web browsers.

Since the Beckman was founded, the interdisciplinary model has spread around the world, countering the trend towards specialization that had dominated science since the Second World War. Cross-cutting institutes have sprouted up in the United States, Europe, Japan, China and Australia, among other places, as researchers seek to solve complex problems such as climate change, sustainability and public-health issues. The interdisciplinary trend can be seen in publication data, where more than one-third of the references in scientific papers now point to other disciplines (see page 306). "The problems in the world are not within-discipline problems," says Sharon Derry, an educational psychologist at the University of North Carolina at Chapel Hill who studies interdisciplinarity. "We have to bring people with different kinds of skills and expertise together. No one has everything that's needed to deal with the issues that we're facing."

Even so, supporters of interdisciplinary research say that it has been slow to catch on, and those who do cross academic disciplines face major challenges when applying for grants, seeking promotions or submitting papers to high-impact journals. In many cases, scientists say, the trend is nothing more than a fashionable label. "There's a huge push to call your work interdisciplinary," says David Wood, a bioengineer at the University of Minnesota in Minneapolis. "But there's still resistance to doing actual interdisciplinary science."

## HIGHLY DISCIPLINED

The idea of dividing academic inquiry into discrete categories dates back to Plato and Aristotle, but by the sixteenth century, Francis Bacon and other philosophers were mourning the fragmentation of knowledge.

One problem lay in the rapid growth of science: there was too much information spread across the disciplines for any one person to handle. Science historian Peter Weingart of Bielefeld University in Germany points to Carl Linnaeus's taxonomic treatise *Systema Naturae* as an example: between its first edition in 1735 and its last in 1768,

the catalogue swelled from 10 pages to 2,300, covering 7,000 species.

In the nineteenth century, the disciplinary boundaries of the modern university started to take root. The disciplines surged in number and power after the Second World War, as nations, particularly the United States, boosted their research support. "It's the moment when universities increased exponentially," says Vincent Larivière, an information scientist at the University of Montreal in Canada. "And the size of the university increased by creating more departments."

Tensions between the United States and the Soviet Union also played a part, says Weingart. The Soviets boasted a research programme geared towards solving societal problems, for example improving agriculture to boost food security. By contrast, US President Dwight Eisenhower argued that basic research should be untethered. "In the field of intellectual exploration, true freedom can and must be practised," he said in a 1959 speech. And although basic research need not necessarily be disciplinary, it does not have the same pressure towards interdisciplinarity as does applied research.

***"WE HAVE TO BRING PEOPLE WITH DIFFERENT KINDS OF SKILLS AND EXPERTISE TOGETHER. NO ONE HAS EVERYTHING THAT'S NEEDED."***

Specialities proliferated as individual disciplines were repeatedly subdivided. Biology was split into botany and zoology, then into evolutionary biology, molecular biology, microbiology, biochemistry, biophysics, bioengineering and more. Late last year, Jerry Jacobs, a sociologist at the University of Pennsylvania in Philadelphia, counted the number of biology-related departments at Michigan State University in East Lansing. There were nearly 40.

From this thicket, the term 'interdisciplinary' emerged. The earliest citation in the *Oxford English Dictionary* dates back to December 1937, in a sociology journal. But even at that time, some believed that the word was already overused. In a report to the US Social Science Research Council in August that year, a sociologist at the University of Chicago in Illinois lumped 'interdisciplinarity' in with other "catch phrases and slogans which were not sufficiently critically examined" (R. Frank *Items* 40, 73–78; 1988).

As an academic movement, interdisciplinarity caught on during the 1970s and has been growing ever since, says Larivière. He credits that rise in part to libraries, which began to stockpile subscriptions and improved researchers' access to journals in alternative fields. A particle physicist could more easily browse biology journals, say. Furthermore, the US focus began to shift from basic research and scientific liberty back to societal problems such as environmental protection, which can rarely be tackled by a single discipline.

The United States was not alone: in 1994, an influential book partially sponsored by the Swedish Council for Planning and Coordination of Research called *The New Production of Knowledge* (Sage) predicted, among other things, an increasingly interdisciplinary future as science seeks to solve socially relevant questions. That book had an impact, says Larivière, particularly in the European



**INTERDISCIPLINARITY**  
A Nature special issue  
[nature.com/inter](http://nature.com/inter)

Union's Fifth Framework funding programme, which ran from 1998 to 2002 and emphasized interdisciplinary, problem-oriented research.

Soon, interdisciplinary institutes began to sprout up around the world, each with its own unique structure and purpose. One of the first, the Santa Fe Institute, founded in 1984, focused on applying advanced mathematics and computational skills to a range of disciplines. Others, such as

over the past decade, including the Academy for Advanced Interdisciplinary Studies at Peking University in Beijing. The NSFC plans to launch further interdisciplinary projects in the coming years, says Yonghe Zheng, deputy director-general of the foundation's Bureau of Science Policy. "China is a developing country," he says. "So the universities and institutes can quickly set up some new centres which reflect the new trend in interdisciplinary research."

Nanyang Technological University in Singapore established its Interdisciplinary Graduate School in 2012; it already has 335 students, out of a total graduate-school population of 2,000. Nanyang's interdisciplinary graduate programme, which bills itself as the first of its kind in Asia, was designed in part to expand the university's fundraising options, says Bo Liedberg, dean of the programme. Because industry is often focused on real-world problems that cross disciplines, an interdisciplinary programme could foster more collaborations with business, he reasons.

That focus on interdisciplinarity as a revenue stream is widespread, says Merlin Crossley, a molecular biologist and dean of the faculty of life sciences at the University of New South Wales in Sydney, Australia. "There is constant pressure on me to make a cross-faculty, cross-institution alliance," he says. "If I want to build a new building, the more allies I have, the easier it is to raise the money." Arizona State University in Tempe saw its federal funding rise by 162% from 2003 to 2012 as it promoted interdisciplinarity across its campus (see *Nature* 514, 292–294; 2014).

Despite this pressure, interdisciplinarity's reach remains modest. For every Nanyang or Durham, there are hundreds of universities that have not embraced significant change. Departmental dividers remain in place — and in power — at most institutions, says Nancy Andreasen, a neuroscientist at the University of Iowa in Iowa City who co-chaired the committee that wrote the National Academies report more than a decade ago. "It has been an enormous disappointment."

## TEAM WORK

For institutions or programmes that have embraced interdisciplinarity, the transition has not always been easy. The most common mistake is underestimating the depth of commitment and personal relationships needed for a successful interdisciplinary project, says Laura Meagher, a consultant based near St Andrews, UK, who coaches interdisciplinary teams. "You see people who think it's not much more than stapling a bunch of CVs to the back of a proposal," she says. "They don't realize that it takes time to build a relationship."

When the push for collaboration comes from the top, some of that focus on personal relationships could be lost — leaving the project to suffer, she says. The UK Energy Research Centre (UKERC) in London, which since 2004 has coordinated and carried out sustainable-energy research, learned how delicate interdisciplinary relationships can be, says Mark Winskel, a social and political scientist at the University of Edinburgh who evaluated the centre's first decade. Its initial five-year phase went well, he says, and culminated in a key publication: *Energy 2050*, which synthesized the institution's results and translated them into recommendations. But the next five-year phase failed to produce a similar achievement.

Winskel surveyed members and found that changes in the UKERC's structure designed to open it to a wider community — for example by offering several rounds of fresh grants

***"THERE IS CONSTANT PRESSURE ON ME TO MAKE A CROSS-FACULTY, CROSS-INSTITUTION ALLIANCE. IF I WANT TO BUILD A NEW BUILDING, THE MORE ALLIES I HAVE, THE EASIER IT IS TO RAISE THE MONEY."***

the Massachusetts Institute of Technology's David H. Koch Institute for Integrative Cancer Research in Cambridge, or the neuroscience-focused Janelia Research Campus in Ashburn, Virginia, tackle questions within a specific discipline but draw in work from other fields. And some, such as the Monash Sustainability Institute in Clayton, Australia, focus on specific problems.

Even as the trend gained momentum, interdisciplinary researchers continued to hit the same hurdles that Brown had encountered. In 1998, chemist Richard Zare at Stanford University in California helped to launch the interdisciplinary institute Bio-X. But an influential colleague urged him not to move his lab into the Bio-X building. Doing so would essentially take Zare away from the chemistry department and his committee and teaching duties there, the colleague argued, weakening the department.

Although he was well established, Zare worried about going against the establishment. "It was very serious," he says. The risk is even greater for young professors seeking tenure, he notes.

In 2004, in response to the growing interest in interdisciplinary work — and the challenges that face those who attempt it — the US National Academies released a report called *Facilitating Interdisciplinary Research*. The authors advised institutions to lower barriers, for example by making budgets flexible so that costs could be shared across departments.

The publication drew a large audience. It has been downloaded more than 7,600 times and had impact beyond US shores. At Durham University, UK, says physicist Tom McLeish, administrators referred to the report when they were forging a series of on-campus interdisciplinary centres. Around that time, McLeish was serving as pro-vice-chancellor of research, and saw interdisciplinarity as a way to make the small university shine on the world stage. He battled with department chairs who feared that the centres would reduce their budgets, and he worked to set up a promotion system that rewards investigators on large team grants in the same way as those on single-investigator grants. The university now has interdisciplinary centres on topics ranging from resilience — both ecological and psychological — to the history of medieval science.

The interdisciplinary trend is also growing in Asia. In 2000, the National Natural Science Foundation of China (NSFC) laid out a plan for interdisciplinary research, and universities have launched several cross-cutting centres



in the middle of phase two — had upset some established long-term relationships. “We became a more diverse community of scholars and disciplines,” he says. “But that also means you become less cohesive.” The UKERC learned from the experience: its third phase, launched in May 2014, aims to provide more stability for collaborative relationships.

Social scientists in particular often face that lack of cohesion, says Thomas Heberlein, a social psychologist at the University of Wisconsin–Madison. When funders emphasize the societal impacts of the work they support, social scientists are often called in to assess the broader implications of a project. But, he says, it is obvious — and insulting — when a social scientist is asked to join a project as a way to tick a box, without a true commitment to incorporating the discipline into the project.

## SOCIAL STRUGGLE

Several UK studies have found that social scientists are less likely than researchers in other disciplines to want to participate in interdisciplinary projects. For Heberlein, who has long collaborated with ecologists and environmental scientists, one of the stumbling blocks is what he calls “the hegemony of the natural sciences”. Those disciplines tend to be held in higher esteem than more qualitative fields such as the social sciences, and they are deemed more rigorous by funders and researchers, he says. That imbalance leads to frustration and undermines collaboration. Heberlein, whose speciality is in conducting surveys of public opinions, says that natural scientists often naively suggest that they can design and execute surveys themselves using an Internet tool such as SurveyMonkey. Heberlein disagrees: “It’s really hard to do the stuff we do,” he says. “Our measurements are complicated.”

Lack of respect can run in many directions when different kinds of researchers come together. Wood says that bio-engineers are always cautioned against having their grants reviewed by panels of biologists, who may be dismissive of engineering research goals and measurements. But he has also served on review panels in which engineers have recoiled at the limitations of clinical research.

As more researchers become involved with interdisciplinary work, the mutual suspicion has started to ease. There have also been some signs of success in the funding arena. The US National Institutes of Health (NIH), for example, says that interdisciplinary proposals fare as well as, or slightly better than, more conventional applications. The European Research Council, by contrast, has noted that interdisciplinary grant proposals on average do not fare as well in review panels as projects that are narrower in scope.

The atmosphere for publishing is also mixed. Interdisciplinary researchers have long complained that it is difficult to get their papers into top-tier disciplinary journals. Heberlein says that the rise of interdisciplinary journals has helped in his field, but he worries about the standard of some of the papers they publish. And he questions the wisdom of training graduate students across disciplines before they have immersed themselves in the rigours of one area. “You’ve got to develop your disciplinary skills first,” he says. “The bad news is the quality of this research is pretty bad and may be getting worse.”

Many view the institutional push for interdisciplinarity as an experiment in progress. “The celebrations have begun, but the actual data on what kind of difference this makes are not in,” says Scott Frickel, a sociologist at Brown

University in Providence, Rhode Island.

As more institutions adopt new ways to organize research, some are also trying to rethink their assessment processes, says McLeisch. In July, he and his colleagues at Durham released a report called *Evaluating Interdisciplinary Research*, and he was surprised when academic societies and funders flocked to learn more. “We didn’t anticipate that we’d be launching this report into an atmosphere where everyone wants to know this,” he says.

And the pace of change varies across the globe. In the United States, the NIH ran a programme to stimulate interdisciplinary research from 2004 to 2012. It resulted in some changes, such as starting to recognize multiple principal investigators on what had been considered single-investigator grants — a switch that removed a disincentive to collaborate. Since then, the agency has not perceived a need to follow up with any other incentives, noting that there are more than 4,000 active NIH-funded research projects that bill themselves as interdisciplinary. “Our general sense is that interdisciplinary research has become a very standard way of doing science,” says Betsy Wilder, head of the NIH Office of Strategic Coordination. “It really pervades NIH funding.”

In some other countries, the experiment has just begun. Chemist Ayyappanpillai Ajayaghosh, director of the National Institute for Interdisciplinary Science and Technology in Thiruvananthapuram, India, says that momentum is building in his country to promote more interdisciplinary projects. In

***“YOU SEE PEOPLE WHO THINK IT’S NOT MUCH MORE THAN STAPLING A BUNCH OF CVS TO THE BACK OF A PROPOSAL. THEY DON’T REALIZE THAT IT TAKES TIME TO BUILD A RELATIONSHIP.”***

Japan, theoretical physicist Tetsuo Hatsuda left the University of Tokyo in part because he felt that the boundaries between disciplines were too heavily enforced there. In 2013, he joined the RIKEN research institute in Wako, Japan, and launched an interdisciplinary team of theoretical physicists, chemists and biologists to work out techniques that will accelerate all three fields. He hopes that the effort will stimulate more interdisciplinary work in the country. “Japan is a little behind other countries,” he says. “Theoretical science is a good starting point because it is easy for us to interact.”

Some 25 years after it opened, the Beckman Institute’s experiment in interdisciplinary research has been a success, says Brown. The centre continues to attract distinguished faculty members and large team grants — last year it won a research contract worth up to \$12.7 million from the federal government’s Intelligence Advanced Research Projects Activity programme — even though competition for such money has increased as more universities build interdisciplinary teams.

And Brown bristles at the suggestion that the global push for interdisciplinarity might be a fad. “The answer is a resounding ‘no,’” he says. “Things have changed — now people focus on big problems, and if you go for a big problem you need to be interdisciplinary.” ■ [SEE EDITORIAL P.289](#)

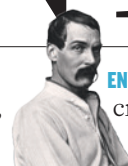
**Heidi Ledford** writes for Nature from Boston, Massachusetts.

# COMMENT

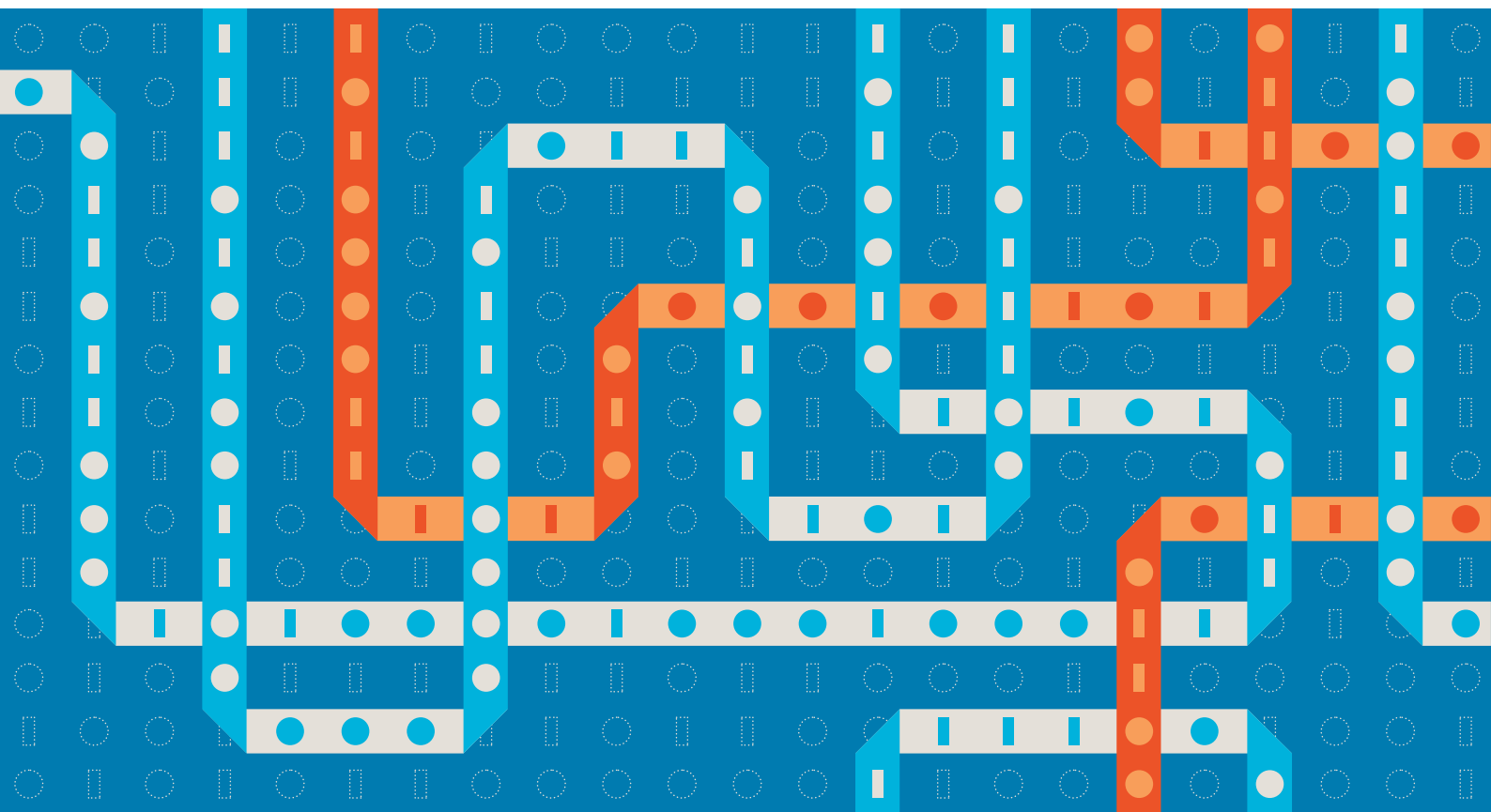
**INTERDISCIPLINARITY** Five principles for fruitful partnerships **p.315**

**INTERDISCIPLINARITY** Scholars and technologists probe texts ancient and modern **p.318**

**HISTORY** Ethnographer Richard Francis Burton, reappraised **p.320**



**ENVIRONMENT** Tally of crackdowns under China's revised law **p.321**



## Global funders to focus on interdisciplinarity

Granting bodies need more data on how much they are spending on work that transcends disciplines, and to what end, explains **Rick Rylance**.

**T**hree arguments are often made in favour of interdisciplinary research. First, complex modern problems such as climate change and resource security are not amenable to single-discipline investigation; they often require many types of expertise across the biological, physical and social disciplines. Second, discoveries are said to be more likely on the boundaries between fields, where the latest techniques, perspectives and

insights can reorient or increase knowledge<sup>1</sup>. The influence of big-data science on many disciplines is a good example. Third, these encounters with others benefit single disciplines, extending their horizons.



**INTERDISCIPLINARITY**  
A *Nature* special issue  
[nature.com/inter](http://nature.com/inter)

The arguments against interdisciplinary work are also familiar. Devotees of normalized citation measures often contend that interdisciplinary research is inferior. Some fear that it drains funds, time and energy from 'core' disciplines. Research funders often hear complaints that schemes targeted at interdisciplinarity distract researchers. There is a persistent argument that 'you can't have inter-disciplines without disciplines.' ►

WES FERNANDES/NATURE



► According to proponents of interdisciplinarity, obstacles abound. Academic institutions' budgets, governance and promotion arrangements are usually organized around single disciplines, as are processes at many granting bodies and journals. Interdisciplinary research struggles for prestige — as measured by quantitative metrics that favour single disciplines — and it is trickier to peer review. Thus early-stage researchers are often advised that starting on an interdisciplinary trajectory is not a smart move.

One striking aspect of this debate is how poor the consolidated data are on which to base judgements. This is why the Global Research Council (GRC) has selected interdisciplinarity as one of its two annual themes for an in-depth report, debate and statement between now and mid-2016. (The other is the position of women in science and research.) The GRC is a federation of more than 50 national research funders, with representatives from countries including Brazil, China, Japan, Russia, the United Kingdom and the United States. Participants include the US National Science Foundation, Research Councils UK (RCUK), Science Europe and the Chinese Academy of Sciences. I serve on the GRC's governing board, in my capacity as chair of RCUK.

As it has done in recent years with peer review and open access, the GRC aims to establish a common position on interdisciplinarity — a topic on many people's minds worldwide, and one in which I have a personal interest.

## GROUND TRUTH

So, what do we know? The 2014 Research Excellence Framework (REF) — a multi-year UK exercise that assessed universities' research strengths in 2008–13, and which thus determines funding — found that, when academics were asked to submit cases of research to REF that had significant impact outside academia, 80% were interdisciplinary. However, items submitted to discipline-based REF panels under-represented the quantity of top interdisciplinary research published by UK researchers in some fields<sup>2</sup>. These included health sciences, mathematics, information technology and the humanities. This is despite growth in UK interdisciplinary work overall. (The United Kingdom's share of the top 10% most interdisciplinary research grew from 7.9% to 9.1% in the four years to 2013.) In my view, this suggests that researchers perceive interdisciplinary research to be vulnerable to discipline-based assessment.

Further evidence comes from the UK government's recent triennial review of the country's seven national research councils<sup>3</sup>. The review heard 'evidence' — what I consider opinion — to the effect that current structures did not serve interdisciplinary

research well, and that it was significantly more difficult to gain funding for this than for mainstream activity. The review recommended that RCUK — the councils' umbrella body — investigate this, which it has been doing.

It is difficult to get clear answers in response to the allegation that funding is more difficult to obtain for interdisciplinary work. Sample tests do not sustain the view that success rates for interdisciplinary grants are significantly adrift. But funding data are not easily analysed in this way. This is in part because there are different schemes under which interdisciplinary work is undertaken: for example, through 'grand challenge'-style programmes, fellowships or 'highlighted' opportunities in mainstream schemes. Awards are also made in areas in

***"The generic protocols of a scientific paper and those for a piece of humanities research are very different."***

which interdisciplinarity is simply the norm, such as design. So, what should be included? More fundamental, however, is an issue of definition. What should be measured when evaluating the

funding of interdisciplinary activities?

Arcane debates about whether research is inter-, multi-, trans-, cross- or post-disciplinary complicate data collection. People also speak of methodological, theoretical, instrumental, critical, restructuring and bridge-building interdisciplinarity<sup>4</sup>. I find this faintly theological hair-splitting unhelpful. But there are areas in which discrimination is important. One is the difference between 'near-neighbour' or 'distant' disciplines.

Interdisciplinary research that involves neighbour disciplines is much more common, and significantly easier to develop, than areas in which the disciplinary stretch is vast and the logistics and intellectual challenge more demanding. This seems a significant point of analysis and one featured in a study<sup>2</sup> by the publisher Elsevier, which used a citation-based approach to review interdisciplinarity in the United Kingdom. The measure considered the diversity of citations and the disciplinary distance between them to determine the extent of a paper's disciplinary reach. The German Research Foundation (DFG) has used similar techniques for its funding portfolio, again demonstrating significant differences between 'near' and 'far' interdisciplinarity — far research being more complex to undertake<sup>5</sup>.

## CASE STUDY

I have personal experience of the challenges of interdisciplinary working. My background is in English literature, but I have worked for many years on the history of psychology, in particular on the intersection of

mind and biomedical systems. Separately, I work with neurologists on what the brain is doing when a person reads complex verbal artefacts such as poems. This is tested experimentally using functional magnetic resonance imaging.

My personal interest is in why, in brain-processing terms, might culture be good for you (if it is)? Clinicians have different — but compatible — concerns, for example in recovering advanced reading functions and well-being following head injury. Educationists are interested in information processing and interpretation.

Of my two areas of research — one historical, the other experimental — the first is not much of a stretch, intellectually or methodologically. The second is. I had to learn new things: to work in a team, to work with complicated machinery, to observe ethical protocols and to raise money. I have had to acquire knowledge of brain anatomy and statistical analysis, and learn a different research mindset. This has been far from straightforward. It has meant, for instance, adjusting how I think about elementary issues such as 'what constitutes sufficient, appropriate evidence?'; methods of analysis; how inferential conclusions can be sustained; and how to write up results.

The generic protocols of a scientific paper and those for a piece of humanities research are very different. This is a matter both of how to express oneself and of the way the proposition is shaped in the first place. I have found that it is easy to be too 'arty' for the scientist and too 'sciencey' for the arts researcher. A humanities colleague remarked that the statistics "might as well be in Russian"; a scientist asked why the poems we used in the neurology experiments were by different people (for example, Shakespeare and Milton): couldn't we just write our own for consistency?

And then there is the question of serial investigation. The cycle of grant, paper, grant, paper and so on does not pertain in the humanities, in which articles tend to emerge from longer projects that culminate in a book. In my experience, issues about raising grants (from whom?), satisfying peer review (from which constituency?) and gaining career recognition are relevant. But paramount is confronting the groundwork challenges that come with interdisciplinary work — especially those that require 'stretch' — and doing so with integrity, honesty and a degree of disciplinary self-denial.

There is evidence that the first steps in establishing interdisciplinary projects are crucial. This was a finding of a review<sup>6</sup> of the European Union's efforts to stimulate interdisciplinary work under its Fifth Framework Programme for research development. Projects did not succeed as well as they might have because they did not facilitate 'enabling

conversations' from the outset and because they lacked coherent leadership. Interdisciplinary work requires particular skills, mindsets and attention to establishing common ground<sup>6–8</sup>.

### FACT FINDING

Interdisciplinarity will be a headline topic at the GRC annual meeting in Delhi in May 2016, organized by India's Science and Engineering Research Board and RCUK. A report on the state of play worldwide is being commissioned by RCUK, on behalf of the GRC (the team to undertake the research will be appointed in October).

The report will survey current policy and practice among global research funders. What forms of support do they offer to interdisciplinary research? How and where is it done? What are its outputs and impacts? The survey will begin to establish base data on how interdisciplinarity can best be stimulated and managed, and look for good practice in this most precious and complex of research endeavours.

The GRC expects to issue a policy statement following this meeting, as it has done previously on topical areas. These documents focus and clarify attitudes on key subjects. They marshal data that can be used while national policies are established and international cooperation is developed. We need much better definitions of what kind of thing we are supporting when and if we support interdisciplinary research, and better intelligence about what works. ■

**Rick Rylance** is chief executive of the Arts and Humanities Research Council, chair of Research Councils UK, and a member of the governing board of the Global Research Council.  
e-mail: r.rylance@ahrc.ac.uk

1. Lakhani, K. R., Jeppesen, L. B., Lohse, P. A. & Panetta, J. A. *The Value of Openness in Scientific Problem Solving* Harvard Business School Working Paper (2007).
2. Elsevier. *A Review of the UK's Interdisciplinary Research Using a Citation-based Approach* (HEFCE, 2015).
3. Department for Business, Innovation & Skills. *Triennial Review of the Research Councils* (BIS, 2014).
4. Klein, J. T. in *The Oxford Handbook of Interdisciplinarity* (eds Frodeman, R. et al.) 15–30 (Oxford Univ. Press, 2010).
5. German Research Foundation. *Interdisciplinary Review Processes: Structural Impact and Funding Success* (DFG, 2013); available at <http://go.nature.com/uyfxlp> (in German).
6. Bruce, A., Lyall, C., Tait, J. & Williams, R. *Futures* **36**, 457–470 (2004).
7. McLeish, T. & Strang, V. *Leading Interdisciplinary Research: Transforming the Academic Landscape* (Leadership Foundation for Higher Education, 2014).
8. Whitfield, J. *Nature* **451**, 872–873 (2008).



Equipping cities to weather our changing climate takes many disciplines working together.

# How to catalyse collaboration

Turn the fraught flirtation between the social and biophysical sciences into fruitful partnerships with these five principles, urge **Rebekah R. Brown, Ana Deletic and Tony H. F. Wong.**

**A**n urgent push to bridge the divide between the biophysical and the social sciences is crucial. It is the only way to drive global sustainable development that delivers social inclusion, environmental sustainability and economic prosperity<sup>1</sup>. Sustainability is the classic 'wicked' problem<sup>2</sup>, characterized by poorly defined requirements, unclear boundaries and contested causes that no single agency or discipline is able to address<sup>3</sup>.

It is crucial to understand, then, why so many well-meaning attempts at interdisciplinary collaboration fail to deliver tangible outcomes — and why others succeed. Here we offer an unapologetically personal answer by reflecting on how, working across multiple faculties of Monash University in Melbourne, Australia, we have built a team of

disciplinary experts that delivers integrated and sustainable water management across multiple cities.

We have now grown this interdisciplinary team to incorporate other institutions nationally and internationally. At the same time, we acknowledge that substantial transaction costs come with interdisciplinary research — it takes extra time and effort to make it work.

### PERSONAL JOURNEY

Our journey began in the early 2000s, with two maturing groups working on urban water research: one in the faculty of engineering, focused on sustainable stormwater technologies, and the other in the faculty of arts, focused on urban water governance (see Supplementary Information; [go.nature.com/pjgbmn](http://go.nature.com/pjgbmn)). The research teams had a common impact agenda, and our collaboration grew from a realization that an interdisciplinary approach would be more effective. In 2005, the two groups joined and secured funding for the establishment of a Aus\$4.5-million ►





► (US\$3.1-million) Facility for Advancing Water Biofiltration<sup>4</sup> that brought together more than 20 Monash researchers and PhD students across civil engineering, ecology and sociology. By 2012, this had culminated in the award of a Aus\$120-million Cooperative Research Centre (CRC) for Water Sensitive Cities. It comprises a partnership of more than 85 organizations, including 13 research institutions, and around 230 researchers and PhD students from more than 20 disciplines and subdisciplines across the social and biophysical sciences and humanities.

Over the past decade, our collaborations have increasingly made a practical difference. We produce regular synthesis documents (see, for example, ref. 5) containing technology information and enabling policy advice, written in an accessible way to facilitate engagement and uptake. These have been heavily used in policy and strategy documents, which speeded up the adoption of our research. For example, stormwater regulations introduced in the state of Victoria in 2006 were underpinned by our research, and other state and local governments in Australia have adopted our recommended performance targets for the management of urban run-off. As a consequence, our stormwater-biofiltration technology has been increasingly adopted in cities across Australia<sup>6</sup>, Singapore, China and Israel. Since 2010, our expanded framework for integrated city-wide water-cycle management<sup>7,8</sup> has been used by governments (such as those of Australia, Singapore and China) and international organizations (such as the

Asian Development Bank) to guide their strategic planning and investment.

In that time, we have had to resolve considerable tension, which hinders meaningful collaboration. The biophysical sciences tend to have well-agreed theories; the social sciences spend much time developing (and often disagreeing on) theoretical questions. Both fields have control and comparison at their core. But biophysical researchers mainly perform quantitative research (often in well-controlled and replicable laboratory conditions), whereas social science can be qualitative or quantitative, and also use interpretative validation approaches.

We witnessed biophysical researchers accusing social scientists of poor rigour and of spending too much time conceptualizing problems without exploring and offering solutions. Conversely, social scientists were often frustrated that biophysical researchers were too focused on solutions, reductively overlooking the wider societal implications of their proposed solutions.

This discord is exacerbated by an inherent cultural hierarchy that often privileges the biophysical over the social sciences. Environmental problems have typically been framed from a biophysical perspective, meaning that social scientists are not effectively engaged in developing integrated solutions<sup>9</sup>.

#### FIVE PRINCIPLES

The journey was not for everyone, and we lost some talent along the way. Yet many stayed on. How did we help academics to overcome these biases? We used these five principles.

**Forge a shared mission.** Driving our collaborative journey was the shared mission of delivering water-management strategies that address the challenges of floods, droughts and degraded waterways. This approach fosters more sustainable, resilient, productive and liveable cities — for a healthy planet and population. The shared mission provided a compelling account of the overall goal of the collaboration, included impact as a necessary outcome, and was sufficiently broad to incorporate meaningful roles for all disciplinary researchers involved.

This mission also maintained a sense of purpose in the face of occasional failure and of the ongoing investment of huge time and effort to appreciate the norms, theories and approaches of other disciplines. When we needed the input of certain disciplines, and hastily included researchers that did not share the mission, it was not a success. The subsequent departure of these researchers from the team initially weakened the skill set of the group, but provided the motivation to expand our collaboration across multiple institutions.

**Develop ‘T-shaped’ researchers.** In our experience, interdisciplinary collaborations have the greatest chance of success when researchers are ‘T-shaped’<sup>10</sup> — able to cultivate both their own discipline, and to look beyond it. Breadth and depth are key. T-shaped researchers build credibility by aiming for the highest scientific contribution in their field — a point of particular importance for early-career researchers, whose prospects

## MAKE IT MAINSTREAM

### *Ways to promote interdisciplinary research*

#### Funders

- Manage funding from an interdisciplinary perspective while reinforcing research impact. Discipline-based agencies must form joint funding programmes.
- Panels should include a balance of experts from the social and biophysical sciences, with a strong appreciation of other disciplines. It is also useful to include end-users of the research (for example, practitioners and policymakers).
- Calls for funding should request balance between disciplines and prefer teams that have a proven record of collaboration. Publication in applicants' own disciplines should be essential; publishing in other disciplines is desirable.

#### Institutions

- Introduce key performance indicators that promote T-shaped researchers. For example, include qualitative measures of impact on policy and practice, as well as conventional academic indices.
- Identify institutional research strengths that show potential for interdisciplinary collaboration and incentivize it through seed grants.
- Reduce transaction costs: for example, through summer schools to develop constructive dialogue skills. Provide platforms — seminars, research workshops, debating competitions — to discuss challenges

in cross-disciplinary research and offer insights into the norms and cultures of other disciplines. Co-locate researchers from different disciplines who work on the same grand challenges.

- Invest in interdisciplinary PhD cohorts, co-supervised by academics from diverse departments or faculties.

#### Publishers

- Invest in and create high-quality interdisciplinary journals, managed by editorial teams or boards of T-shaped researchers.
- Run special issues in high-impact, single-discipline journals that focus on interdisciplinary research.
- Peer reviewers should assess work using their disciplinary expertise, while being tasked to be open to innovations across disciplines.

#### Researchers

- Build stamina, patience and self-awareness to manage the long journey of establishing a productive interdisciplinary team.
- Put your best ideas forward even if they are unfinished, and be open to alternative perspectives from other disciplines, policymakers, industry practitioners and community members.
- Prioritize depth early on, and embrace breadth by building relationships with those from other fields and practices.

for promotion are judged against research excellence criteria (see principle 5). T-shaped researchers also engage actively with other disciplines (see principle 3) to understand and appreciate their norms, theories, approaches and breakthroughs.

Many believe that interdisciplinary research delays career progression or is the luxury of senior researchers. This has not been our experience: many of our researchers were able to maintain a high publication rate in their own discipline, and — as part of a team — secure increasing interdisciplinary research funding. However, it took nearly five years to start publishing our joint interdisciplinary research in high-impact journals.

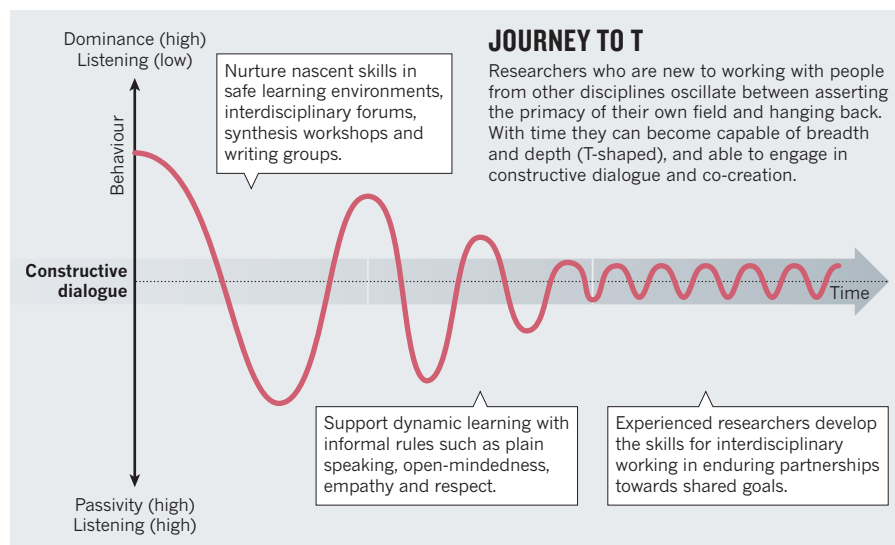
**Nurture constructive dialogue.** Through a decade of trial and error, we have invested heavily in creating the environment and informal rules that empower researchers across all sciences to engage effectively, despite their vastly different approaches to research design and methodology, and their differing technical vocabularies and communication cultures.

This has involved some commitments: to interact in plain English (disciplinary jargon is frowned on); to foster empathy and respect for different disciplinary norms; and to reflect on what is working in collaborative interactions. We designed regular interdisciplinary forums using these rules. This led to the co-development of key publications — for example, through interdisciplinary workshops, we have jointly written three annual reports for policymakers and water practitioners<sup>5</sup>. These activities grew into a sought-after annual short course and a massive open online course (MOOC) showcasing different disciplinary approaches to urban water challenges.

Reaching the ideal of constructive communication across the sciences takes time and practice — researchers new to the group may not yet have the necessary skills. Typically, they pass through three stages of development (see 'Journey to T'). Initially, new collaborators tend to dominate discussions and assert the primacy of their discipline. Soon after, they recognize the importance of other disciplines and adopt a more passive demeanour. Eventually, the researchers settle into a space of constructive dialogue.

We find that some quit and others stay to become mature collaborators, able to co-create across academic disciplines and broader networks. The role of more experienced collaborators is to support new colleagues' personal journeys into these dynamic relationships.

**Give institutional support.** Academic career pathways for interdisciplinary research are essential if it is to attract and



retain the brightest and best. Monash University's senior leadership team consistently signalled that it values research that is interdisciplinary, attracts significant industry involvement and delivers real-world impact — despite the organizational structures and global academic norms that are biased towards more conventional, disciplinary approaches.

This value was communicated to researchers through university policies, promotion criteria and seed-funding programmes. For example, the engineering faculty has introduced qualitative research standards (alongside the conventional quantitative

*“Despite our rewarding experience, interdisciplinary research is still on the margins.”*

measures), that attempt to measure the impact of research on practice. The faculties of engineering and arts now award small competitive grants to teams from both facul-

ties to catalyse collaborations.

Monash has established a PhD programme for cohorts of students working on a common global challenge across a number of disciplines; for instance, sustainable urban water management in developing Asian cities. These groups work in a constructive dialogue environment.

**Bridge research, policy and practice.** Finally, the establishment of enduring connections between researchers, policymakers and industry practitioners proved to be an important driver in growing our interdisciplinary collaborations. Refreshingly, industry rarely thinks in disciplinary silos. They tend to tackle complex problems from a range of perspectives, thereby modeling integrated, solution-focused thinking.

To ensure real-world impact, we engaged

policy and industry partners in the design of our research programme and encouraged them to critique our scientific approach and presentation of results. We also ran frequent events that allowed professionals from policy and industry to interact with researchers. For example, in 2008, through a national roadshow, we showcased how our research is addressing crucial water challenges around Australian cities. Aimed at policymakers and industry and community leaders, it stimulated research and partnerships.

Despite our rewarding experience, interdisciplinary research is still on the margins. We urge researchers, institutions, and funding bodies committed to sustainable development to make it mainstream (see 'Ways to promote interdisciplinary research'). ■

**Rebekah R. Brown, Ana Deletic and Tony H. F. Wong** are at Monash University in Melbourne, Australia, and in the Cooperative Research Centre for Water Sensitive Cities. R.R.B. is also director of the Monash Sustainability Institute.  
e-mail: rebekah.brown@monash.edu

1. United Nations. *Transforming our World: The 2030 Agenda for Sustainable Development* (UN, 2015).
2. Rittel, H. W. J. & Weber, M. M. *Policy Sci.* **4**, 155–169 (1973).
3. APSC. *Tackling Wicked Problems: A Public Policy Perspective* (Australian Government, 2007).
4. Deletic, A., Fletcher, T. D., Brown, R. R., Hatt, B. E. & Wong, T. H. F. *Water* **35**, 64–72 (2008).
5. Wong, T. H. F. et al. *blueprint2013 — Stormwater Management in a Water Sensitive City* (Cooperative Research Centre for Water Sensitive Cities, 2013).
6. Brown, R. R., Farrelly, M. A. & Loorbach, D. A. *Glob. Environ. Change* **23**, 701–718 (2013).
7. Wong, T. H. F. & Brown, R. R. *Water Sci. Technol.* **60**, 673–682 (2009).
8. Brown, R. R., Keath, N. & Wong, T. H. F. *Water Sci. Technol.* **59**, 847–855 (2009).
9. ICSU. *Earth System Science for Global Sustainability: The Grand Challenges* (International Council for Science, 2010).
10. Hansen, M. & von Oettinger, B. *Harvard Bus. Rev.* **79**, 106–116 (2001).



## INTERDISCIPLINARITY

# Inside Manchester's 'arts lab'

**Peter E. Pormann** on the revelations a meshing of technology and humanities can yield.

Digital pioneer Steve Jobs delivered a potent commencement address at Stanford University, California, in 2005. He described how, as an undergraduate, he had studied calligraphy rather than his prescribed curriculum (he later dropped out). Calligraphy may have seemed at the time to have no practical application, but a decade later, when Jobs was working on the Mac, it enabled him to promote proportional fonts and establish Apple as the gold standard in desktop publishing. Jobs fruitfully combined the “liberal arts and technology” — a phrase he used repeatedly in his last keynote addresses before his death in 2011.

Productive interaction between the arts and sciences is at the heart of the John Rylands Research Institute at the University of Manchester, UK. Founded in April 2013, the institute (which I direct with associate director and head of special collections Rachel Beckett) now has a staff of more than two dozen. It brings together scientists, conservators, curators, digital-imaging specialists and humanities scholars to unravel, reveal and realize the research potential of the University of Manchester Library's special collections. These run from clay tablets to e-mail archives. Highlights include Greek, Coptic and Arabic papyri, medieval Hebrew and Persian manuscripts and early-modern printed books — such as one of the world's finest collections of volumes printed by Renaissance humanist Aldus Manutius. The institute was established in response to the rise of digital humanities, a field that enables the study of books and manuscripts in ways that were unimaginable a generation ago.

There have been triumphs and tribulations. We have raised more than £3 million (US\$4.6 million) in funding from sources such as the British Academy and biomedical-research charity the Wellcome Trust. The institute sits in the already-crowded John Rylands Library, where its rapid growth is a challenge. But our 'arts lab' is taking research into uncharted territories by shattering disciplinary and institutional divisions.

To make complex collaborations work, we instigated a buddy system. All researchers — PhD students, postdocs, visiting academics and colleagues with funding for pilot



Erased text in the Syriac Galen Palimpsest is made visible by multispectral-image analysis.

studies — are allocated a curator with intimate knowledge of the materials they study. Art-history postdoc Elizabeth Savage, for instance, won a three-year early-career fellowship from the British Academy to study thousands of fifteenth- and sixteenth-century prints collected by Hiero von Holtorp, a nineteenth-century scholar of early printing technology and aesthetics. Her buddy is visual-collections manager Stella Halkyard, who helped to rediscover this remarkable legacy. Savage also works with colleagues at the library's Centre for Heritage

Imaging and Collection Care (CHICC), who pioneer innovations in colour print photography, such as lighting techniques for imaging gold. Combined with close-ups of pigments, these techniques have helped Savage to identify some of the earliest examples of printed gold ink.

Work at the CHICC is also revolutionizing understanding of papyri and palimpsests — manuscripts from which text has been erased to allow reuse of the page. Researchers have made detailed images of artefacts using cutting-edge technology: a 60-million-pixel



**INTERDISCIPLINARITY**  
A Nature special issue  
[nature.com/inter](http://nature.com/inter)

MIKE TOTTH/SIAM BHAYRO/DOUG EMERY/DIGITALGALILEO.NET/CC BY 3.0

digital sensor, combined with a MegaVision EV LED illumination system. This combines high-resolution photography with multispectral imaging, which captures data at frequencies across the electromagnetic spectrum. It can reveal once-unreadable texts, because different inks reflect light in different spectra differently. Thus papyrologist Roberta Mazza has discovered the 'Last Supper amulet', a papyrus with biblical passages on one side and a grain-tax receipt on the other. Mazza traced its provenance to near ancient Hermoupolis in Egypt, close to modern Al Ashmunayn.

We are also collaborating with scientists including Mark Dickinson, a physicist and medical-imaging specialist at Manchester's Photon Science Institute. Medical imaging is rich in techniques that can be used to analyse artefacts, such as optical coherence tomography, which is usually harnessed for imaging tissue or visualizing blood flow. Dickinson has tested it on carbonized papyri too delicate to unroll, revealing hidden text.

Also key to investigating the collections is image analysis. We are using statistical techniques such as canonical variate analysis (CVA), which compares group structures in multivariate data, to read erased text on palimpsests. CVA is applied to a multispectral image and an algorithm is trained to recognize overlying text, the erased underlying text and areas where the two coincide. This effectively maximizes the contrast, so the under-text 'pops' out and becomes more readable.

A £1-million image-analysis project that grew partly out of a collaboration with the

CHICC and has received funding from the UK Arts and Humanities Research Council is studying the Syriac Galen Palimpsest. This is an eleventh-century liturgical work that carries an erased sixth-century undertext — a Syriac translation of *On Simple Drugs* by the classical physician Galen (around AD 129–216). We already had a large data set of multispectral images; now images of the same page are being combined to make the under-text more legible (see picture). Overseeing this is computational primatologist Bill Sellers, who ordinarily uses computer modelling to reconstruct the movements and evolution of extinct species.

All of this work generates large sets of images, stored as TIFF files. These raise the question of how to store and analyse big data. A challenge will be establishing integrated systems to allow comparative research across platforms. For Greek papyri and Hebrew and Persian manuscripts, we plan to develop solutions with the Cambridge Digital Library; this will feed into the iLibrary strategy to bring our digital collections and projects under one roof. We can also look at large amounts of texts and metadata with the tools of computational corpus linguistics — which studies language through samples of real text — and text mining, which hunts through text to extract data. One such tool is the language-processing software system U-Compare.

*"The nature of the institute binds ancient artefacts to state-of-the-art science."*

Some of our collections are born digital — for example, we hold the e-mail archives of local literary publishing house Carcanet — and future researchers will undoubtedly approach these differently from how they look at hand-written correspondence. We have begun to collaborate with computational linguists at Manchester's National Centre for Text Mining, as well as colleagues at the nearby Centre for Translation and Intercultural Studies, who have vast experience with large sets of multilingual texts. And with palaeography — the study of ancient handwritings, their dating and their classification — artificial intelligence might offer research avenues that the institute is keen to explore. By training software to recognize certain hands and writing styles, one might be able to query vast virtual collections of manuscripts in unprecedented ways.

Delivering the institute's inaugural lecture, historian Ann Blair of Harvard University in Cambridge, Massachusetts, said: "In embracing new media, we must never discard the old ones." The interdisciplinary nature of the institute is its signature, the tie that binds ancient artefacts to state-of-the-art science. These form a dual legacy for future generations, who will want to ask different questions of the library's remarkable holdings. ■

**Peter E. Pormann** is founding director of the John Rylands Research Institute at the University of Manchester, UK, and principal investigator on the Syriac Galen Palimpsest project.  
e-mail: [peter.pormann@manchester.ac.uk](mailto:peter.pormann@manchester.ac.uk)

## ANTHROPOLOGY

# One-man multidisciplinarian

Clare Pettitt reassesses the legacy of Victorian polymath Richard Francis Burton.

**R**ichard Francis Burton (1821–90) thirsted for and mastered knowledge in so many fields — from geography to sexology — that his real legacy for science is muddled. The flamboyant polymath was an eminent explorer, a pioneer of ethnography and a linguist fluent in more than 25 languages (from Arabic to Swahili) and a number of dialects. He wrote or translated more than 40 volumes, including *The Lake Regions of Central Africa*, published 155 years ago, and the first English edition of *The Arabian Nights* (1885). He was also an enthusiastic amateur of botany, geology and zoology, even running an experiment on monkey communication while living in Sindh (now Pakistan). Overall, this furiously energetic multidisciplinarian both

contributed vastly to knowledge of other cultures and continents, and sometimes misread them to his — and their — cost.

These complex interests were the fruit of a turbulent mind. The eldest son of an army family, Burton had a protean character shaped on the road as his parents moved their young family restlessly around France and Italy. He started to learn Latin at three years old and Greek at four, and quickly picked up French, Italian and local dialects. At the University of Oxford, UK, contemptuous of the teaching methods, he honed his mastery of languages but was expelled for attending a steeplechase. He was soon propelled into the Bombay Infantry and immersed himself in Indian languages and culture. Violent and mesmerizing by turns, he was viewed as both

prodigiously gifted and morally suspect by his contemporaries — as an 'other', just as he himself was possessed by otherness.

By 1853, Burton had turned to exploration. Still beset by inner conflicts, he could also attract conflict with others. His great 1856–59 expedition to East Africa with John Hanning Speke, instigated by the Royal Geographical Society in London, was a case in point. It made "formidable contributions to imperial knowledge production", according to historian Adrian Wisnicki. Although both men were seriously disabled by disease, Burton became the first European to see Lake Tanganyika. He kept dense geographical and cultural notes and meteorological records, and collected specimens for what are now the Royal Botanic Gardens, Kew, and the ▶



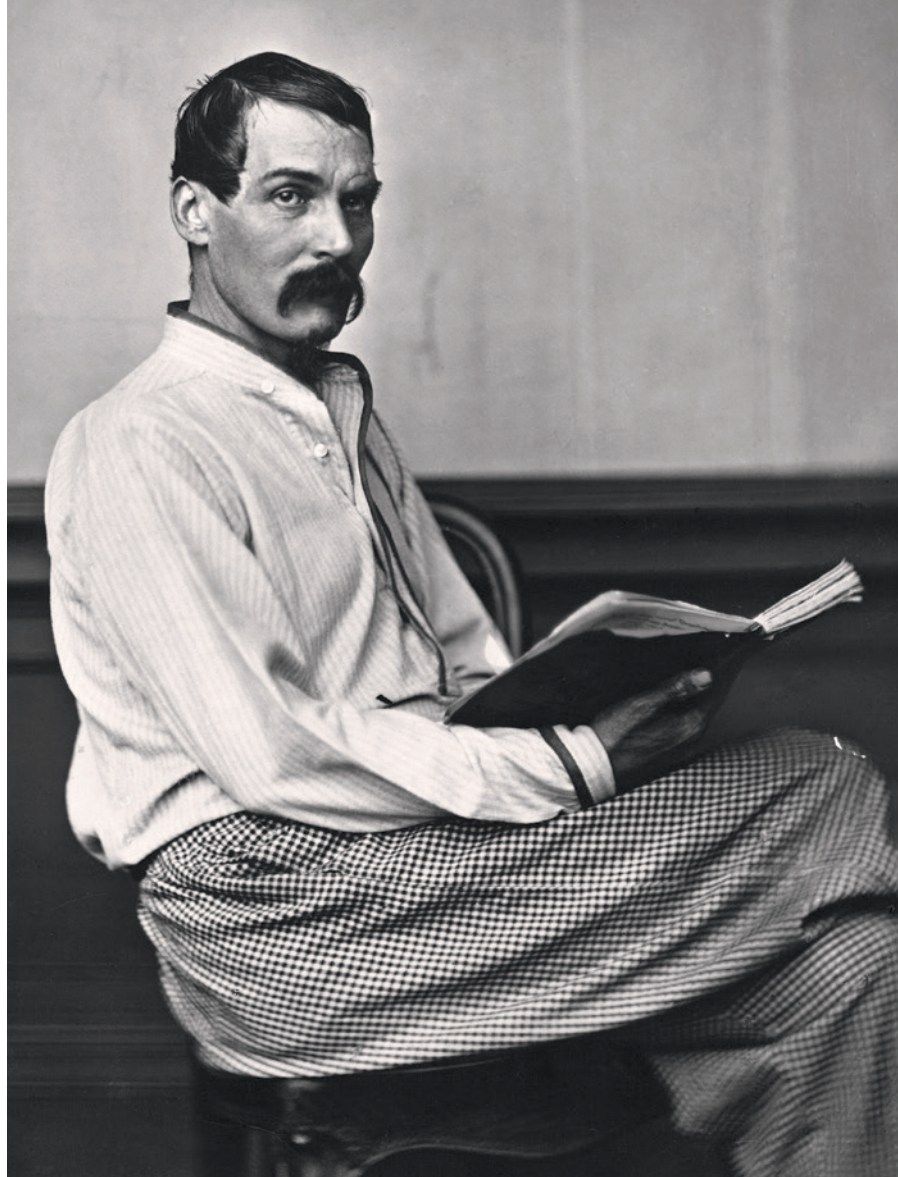
► Royal School of Mines in London. But the expedition led to a bitter rivalry between the two over the source of the Nile, with Speke claiming it as the lake that he dubbed Lake Victoria, and Burton feeling that the evidence failed to add up. Long after their return, in 1864, the British Association for the Advancement of Science called for a debate in London, but Speke died of an unexplained gunshot wound the day before. “The charitable say that he shot himself, the uncharitable say that I shot him,” Burton wrote to a friend.

Burton was shocked, but published *The Nile Basin* that year, reiterating his position in the Nile controversy first detailed in *The Lake Regions of Central Africa*. Burton felt that Speke’s account, *Journal of the Discovery of the Source of the Nile* (1863) had dressed Africa

**“Burton’s immersion in a multitude of languages and cultures gave him a unique perspective on humanity.”**

up in flowery, fundamentally unscientific rhetoric, claiming for instance that a mass of dirty huts (in Burton’s words) was a village built on the most luxurious principles. Burton insisted on using indigenous names and learnt local languages so that he could communicate directly with people he met — and his investigations would prove invaluable to future explorers. “I undertook the history and the ethnography, the languages, and the peculiarities of the people,” he is quoted as saying, adding scornfully that to Speke “fell the arduous task of delineating an exact topography”. Geography, Burton established, was a social as well as a physical science. The explorer Henry Morton Stanley would prove in 1875 that Speke had correctly identified the source of the Nile, but he used Burton’s notes to get there. As Burton put it in *Zanzibar; City, Island, and Coast* (1872), future expeditions “had only to tread in my steps”.

Throughout a life of trailblazing travel and diplomacy — from Somaliland to Benin, Arabia, the Middle East, Asia and the Americas — Burton’s first epistemological framework for colonial encounters was the ‘Orientalist’ one of linguistic scholarship. But as an ethnographer, he was original. He mingled with the people whose cultures he studied, understanding that knowledge is embodied and must be historically contextualized. This was criticized in Victorian England, with its horror of ‘going native’, but places him ahead of his time. Burton was always quick to acknowledge the contingencies and accidents that brought him into contact with local people, and never tried to efface himself from his narrative. Only in the late twentieth century did anthropologists such as John and Jean Comaroff suggest that the obvious weaknesses of ethnography as a ‘science’ are also its strengths, as ‘participant



HULTON-DEUTSCH COLLECTION/CORBIS

Ethnographic pioneer and explorer Richard Francis Burton, photographed around 1860.

observation... connotes the inseparability of knowledge from its knower”. Studies from the 1970s onwards supported this view, including Annette Weiner’s *The Trobrianders of Papua New Guinea* (Holt, Rinehart and Winston, 1988), a reappraisal of Bronislaw Malinowski’s study of the Pacific Trobriand Islands, *Argonauts of the Western Pacific* (Routledge and Kegan Paul, 1922).

In other ways, and much less attractively, Burton was very much of his time. His respect for Muslim culture did not preclude his succumbing temporarily to a vicious racism that became particularly extreme in the 1860s and cannot be exonerated. By the mid-1860s he had become one of Britain’s foremost promulgators of the polygenist thesis that Africans constituted a distinct and inferior species, and he helped to found the Anthropological Society of London, established after a dispute with the monogenist Ethnological Society. By his last decade, Burton had come to his senses,

embracing the view that all of civilization came from Africa, and felt that “negroes... have shown themselves fully equal in intellect and capacity to the white races of Europe and America”. But the damage had been done.

Despite this sorry chapter, Burton’s immersion in a multitude of languages and cultures gave him a unique perspective on humanity, with “the enormous advantage of being capable of comparing native with foreign ideas and views of the world”. He knew that other cultures could never be fully ‘translated’ or subsumed into English, and that this militated against the ethos of Empire. He was perhaps less Orientalist than comparativist and relativist. His contribution to the fledgling social sciences was all the more powerful, perhaps, for having been fed by so many streams of knowledge, even if this makes it less visible to us today. ■

**Clare Pettitt** is professor of nineteenth-century literature and culture at King’s College London. She is the author of *Dr Livingstone, I Presume?* and many articles about exploration and travel in Victorian print culture. e-mail: [clare.pettitt@kcl.ac.uk](mailto:clare.pettitt@kcl.ac.uk)



**INTERDISCIPLINARITY**  
A Nature special issue  
[nature.com/inter](http://nature.com/inter)

# Correspondence

## New environment law shows its fangs

China's revised Environmental Protection Law went into effect on 1 January this year. Severe punishments for polluting businesses swiftly followed.

Some 292 cases incurred an accumulating daily fine within the first 6 months, totalling 236 million yuan (US\$37 million). The highest single levy was 15.8 million yuan (data from the Ministry of Environmental Protection; see [www.mep.gov.cn](http://www.mep.gov.cn)). Over the same period, production was curtailed in 1,092 cases and equipment was locked down in 1,814 instances. Criminal charges were brought against 740 polluting businesses, and 782 were punished with police administrative detention.

Local governments are cooperating with the new law, contrary to earlier misgivings (see B. Zhang and C. Cao *Nature* **517**, 433–434; 2015 and H. Yang *et al. Science* **347**, 834–835; 2015). In Linyi in Shandong province, for example, several dozen businesses (including some responsible for high employment and large tax revenues) have been closed down.

**Dasheng Liu** *Shandong Institute of Environmental Science, Jinan, China.*  
[liu\\_sdiep@126.com](mailto:liu_sdiep@126.com)

## Tailor checklists to clinical teams

The problems of replicating the effects of patient-safety checklist trials in routine practice could be mitigated by adapting checklists for individual hospital environments and teams (see *Nature* **523**, 516–518; 2015). An F-16 fighter aircraft would not rely on a checklist devised for flying a jumbo jet.

For instance, much of the World Health Organization's surgical safety checklist is irrelevant to a cardiac catheterization procedure. There is no general anaesthetic or expected blood loss, for example,

but monitoring kidney function is crucial. We therefore designed a bespoke safety checklist to brief the cardiac clinical team on the planned procedure and on any potential problems. Endorsed by the British Cardiovascular Society ([www.bcs.com/checklist](http://www.bcs.com/checklist)), the checklist is regularly modified in response to end-user evaluation.

Smart electronic checklists will further improve safety by highlighting patient-specific risks and acting as a guide in emergencies and for auditing near-misses.

**Thomas J. Cahill** *Oxford University Hospitals NHS Trust, Oxford, UK.*

**Rod Stables** *Liverpool Heart and Chest Hospital, Liverpool, UK.*  
[thomas.cahill@cardiov.ox.ac.uk](mailto:thomas.cahill@cardiov.ox.ac.uk)

## Mining shell waste will not be easy

If the chemical industry is to profit from refining waste crustacean shells and other by-products of seafood processing, collection problems and food-safety issues need to be overcome (see N. Yan and X. Chen *Nature* **524**, 155–157; 2015).

Gathering sufficient animal feedstock for commercial purposes will be a formidable challenge (R. L. Naylor *et al. Proc. Natl Acad. Sci. USA* **106**, 15103–15110; 2009). The transport and storage of seafood by-products from different processing plants is also likely to be extremely costly.

Moreover, expensive energy-intensive drying of crustacean shells would be necessary to prevent microbial growth and production of carcinogenic bacterial aflatoxins. Other health risks could arise from bioaccumulation of contaminants (such as heavy metals in shells) or from cross-species transmission of pathogens and perhaps even of prions through the food chain (L. Cao *et al. Science* **347**, 133–135; 2015).

**Hong-Wei Xiao, Zhen-Jiang Gao**

*China Agricultural University, Beijing, China.*

**A. S. Mujumdar** *McGill University, Quebec, Canada.*  
[xhwcaugxy@163.com](mailto:xhwcaugxy@163.com)

## Seal of approval for ocean observations

We announce that the Pacific Islands Ocean Observing System was certified last month as the first regional partner to attain full membership of the US Integrated Ocean Observing System (IOOS). This certification is a hallmark of the quality of data provided by the IOOS, to the benefit of the public, the private sector and individuals.

It is also an indicator to the global community that IOOS regional partners providing data from the oceans, Great Lakes and coasts of North America have met rigorous criteria for system oversight, information security, public engagement and financial controls.

The IOOS includes federal and non-federal partners in an interagency investment by the US government of more than US\$2 billion annually for the collection and provision of ocean data and for improved forecast capabilities. It comprises about 10,000 unique oceanographic data sets and some 4,000 services that provide data, metadata and refined data products to tens of millions of US users. For instance, IOOS data are used in search-and-rescue operations and to ensure safe operation of commercial vessels.

Certified IOOS data can be entered in the permanent US archive at the National Centers for Environmental Information and can be used internationally by the Global Telecommunication System for meteorological data.

**Chris E. Ostrander** *University of Hawaii at Manoa, Honolulu, Hawaii, USA.*

**Conrad C. Lautenbacher** *GeoOptics, Dunwoody, Georgia, USA.*  
[chriso@hawaii.edu](mailto:chriso@hawaii.edu)

## Lack of help stymied community care

John Foot's book on psychiatrist Franco Basaglia's movement to reform Italy's psychiatric hospitals ends with the passing of Law 180 in 1978 to close down asylums (see A. Tone *Nature* **524**, 290; 2015). Sadly, the law was poorly implemented owing to woefully inadequate resources.

Families received little or no support in caring for those who returned home. For some it was too much, forcing general hospitals to take up the slack. Psychiatrists found their hands tied when confronted with people who were seriously mentally ill, so many ended up in prison stigmatized as criminals.

Even Basaglia's widow, Franca Ongaro Basaglia — a core member of the reform movement and later an Italian senator — described Law 180 as a failure.

**Laura Spinney** *Paris, France.*  
[lfspinney@gmail.com](mailto:lfspinney@gmail.com)

## Education reforms ring true 50 years on

Stephen Bradforth and colleagues' discussion of what is needed to develop "a science-literate population" (*Nature* **523**, 282–284; 2015) echoes the words of a *Nature* editorial 50 years ago, entitled 'New thinking in undergraduate teaching' (*Nature* **205**, 835; 1965).

According to the editorial, "the student is in danger of spending too much of his [*sic*] limited time memorizing facts, and has insufficient time at his disposal to master the principles underlying his subject and to develop his powers of thought". It continues: "the most important purpose of a university education is to teach the student to think for himself... it may on occasion demand a re-examination of the whole approach to a subject in undergraduate courses." Indeed.

**Barry S. Winkler** *Eye Research Institute, Oakland University, Rochester, Michigan, USA.*  
[winkler@oakland.edu](mailto:winkler@oakland.edu)



# Decompensated cirrhosis and microbiome interpretation

ARISING FROM N. Qin *et al. Nature* 513, 59–64 (2014); doi:10.1038/nature13568

The diagnosis of cirrhosis, especially in the advanced/decompensated stages, is made using simple and inexpensive clinico-radiologic-pathological techniques<sup>1</sup>. Qin *et al.*<sup>2</sup>, whose paper has replicated prior studies<sup>3–5</sup>, reported a relatively novel profile to diagnose cirrhosis using complex stool metagenomics despite having a majority (65% discovery and 76% validation cohorts) decompensated cirrhotic population. We have found that the decompensated cirrhosis cohort, which does not require these complicated diagnostic strategies, was responsible for a significant proportion of these microbiota changes on further analysis of their metagenomics data and using a new cohort of 360 subjects. Therefore, given several confounders and the ease of decompensated cirrhosis diagnosis using current techniques, a careful re-interpretation of newer microbiota-based diagnostic strategies that do not a priori differentiate between early (compensated) and decompensated cirrhosis and treat all people with cirrhosis as one uniform population should be performed. There is a Reply to this Brief Communication Arising by Qin, N. *et al. Nature* 525, <http://dx.doi.org/10.1038/nature14852> (2015).

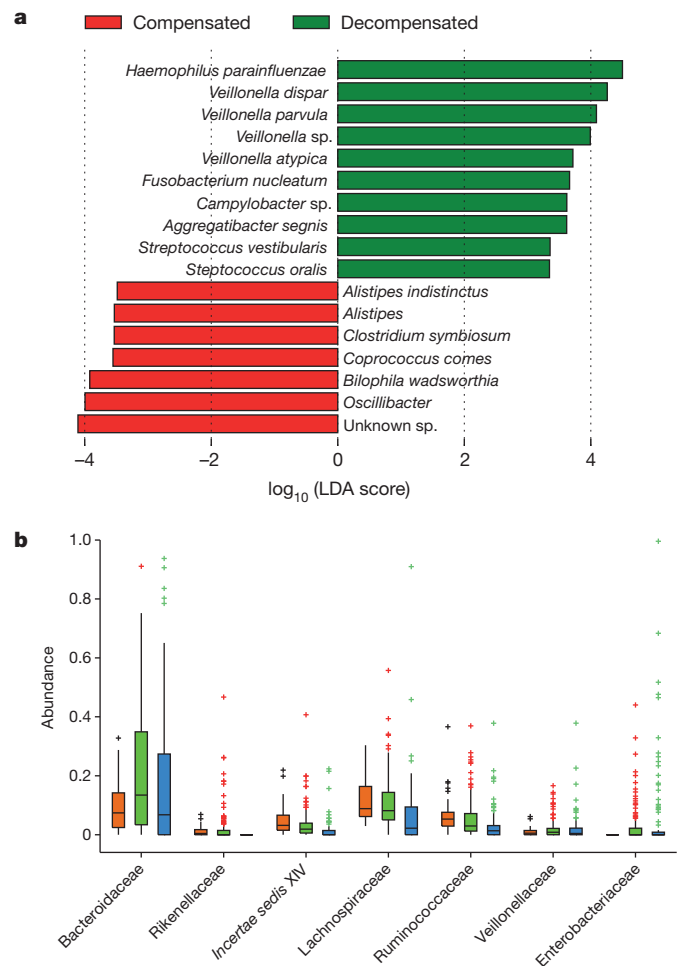
A major confounder in people with cirrhosis are standard of care therapies such as lactulose, rifaximin, antibiotics and acid-suppressants that can affect the gut milieu<sup>1,6</sup>. These alone could explain a large portion of the metagenomics changes and have not been accounted for<sup>5,7–9</sup>. These medications, especially proton pump inhibitors, could also be a major reason why oral origin bacteria are found in the intestine, as has been shown in prospective cirrhotic and non-cirrhotic studies<sup>10,11</sup>.

We hypothesized that there was a significant difference in compensated versus decompensated cirrhotic microbiota in Qin *et al.*<sup>2</sup>, which needs to be accounted for in the interpretation. Using 66 enriched/depleted metagenomic sequences (MGS) provided by S. D. Ehrlich, we performed linear discriminant analysis (LDA) effect size (LEfSe)<sup>12</sup> after classifying them into healthy, compensated and decompensated subjects. LEfSe uses a factorial Kruskal–Wallis and LDA test to detect features with significant differential abundance. We found that even in the selected data set the authors provided, 17 of 66 MGS were different between compensated and decompensated groups (10 MGS overexpressed and 7 MGS underexpressed, Fig. 1a). These included several oral origin species (*Streptococcus oralis* and several *Veillonella* spp.), which were the primary study results. We then enrolled 360 age-matched subjects (45 healthy individuals (age  $54 \pm 3$  years, no chronic diseases), 171 compensated (age  $54 \pm 4$  years, median Child–Pugh score 6) and 141 decompensated cirrhotic patients (age  $55 \pm 2$  years, median Child–Pugh score 9)) for stool multi-tagged pyrosequencing (MTPS)<sup>13</sup>. Using Kruskal–Wallis analysis of relative microbial family abundance  $>1\%$ , we found that compensated and decompensated patients were significantly different (Fig. 1b). Proteobacteria levels, specifically *Enterobacteriaceae*, were significantly higher in decompensated cirrhotic patients. This pattern is also seen in other recent MTPS studies<sup>4,14</sup>. Although MGS and MTPS are not completely comparable, it is interesting that both resulted in similar conclusions. Therefore, there are significant microbiota differences between compensated and decompensated patients that need to be separated in cirrhosis microbial studies.

In addition, in Qin *et al.*<sup>2</sup> the calculation of the model for end-stage liver disease (MELD) score in Supplementary Table 1 is inaccurate, casting doubt on figure 2. The authors compared diabetes patients

with cirrhotic patients to inform their cirrhosis-associated profile. However, diabetes is prevalent and is associated with a poor prognosis in cirrhosis<sup>15</sup>. Therefore these results are not generalizable to patients with cirrhosis and diabetes.

The present need is not for complicated profiles that are unlikely to supplant currently available simple diagnostic strategies, but rather for improving prognostication. This is because gut microbiota are associated with several cirrhosis-related pre-terminal events such as hepatic encephalopathy and infections<sup>1</sup>. A prior study has shown that altered stool microbiota can predict poor outcomes, but further work is required<sup>8</sup>.



**Figure 1 | Microbiota distribution between compensated and decompensated cirrhotic subjects.** **a**, LFSe plot showing metagenomic species that are overexpressed (green) and under-expressed (red) in decompensated compared to compensated cirrhosis from Qin *et al.*<sup>2</sup>. **b**, In the new data set using MTPS, boxplots showing interquartile range of median abundance of statistically significant comparisons between controls (orange), compensated cirrhosis (green) and decompensated cirrhosis (blue) using multiple corrections-adjusted Kruskal–Wallis tests at the family level. The line in the centre shows median.

Therefore, the careful separation of the two groups within cirrhosis, which have different diagnostic criteria and prognoses, and the control of confounders owing to drugs mentioned above, are important for the correct interpretation of these results and to avoid epiphenomena.

**Jasmohan S. Bajaj<sup>1</sup>, Naga S. Betrapally<sup>2</sup> & Patrick M. Gillevet<sup>2</sup>**

<sup>1</sup>Division of Gastroenterology, Hepatology and Nutrition, Virginia Commonwealth University and McGuire VA Medical Center, Richmond, Virginia 23249, USA.

email: jsbajaj@vcu.edu

<sup>2</sup>Microbiome Analysis Center, George Mason University, Manassas, Virginia 20110, USA.

Received 12 October 2014; accepted 8 June 2015.

- Schuppan, D. & Afdhal, N. H. Liver cirrhosis. *Lancet* **371**, 838–851 (2008).
- Qin, N. *et al.* Alterations of the human gut microbiome in liver cirrhosis. *Nature* **513**, 59–64 (2014).
- Bajaj, J. S. *et al.* Linkage of gut microbiome with cognition in hepatic encephalopathy. *Am. J. Physiol. Gastrointest. Liver Physiol.* **302**, G168–G175 (2012).
- Chen, Y. *et al.* Characterization of fecal microbial communities in patients with liver cirrhosis. *Hepatology* **54**, 562–572 (2011).
- Bajaj, J. S. *et al.* Colonic mucosal microbiome differs from stool microbiome in cirrhosis and hepatic encephalopathy and is linked to cognition and inflammation. *Am. J. Physiol. Gastrointest. Liver Physiol.* **303**, G675–G685 (2012).
- Chavez-Tapia, N. C., Tellez-Avila, F. I., Garcia-Leiva, J. & Valdovinos, M. A. Use and overuse of proton pump inhibitors in cirrhotic patients. *Med. Sci. Monit.* **14**, CR468–CR472 (2008).

- Bajaj, J. S. *et al.* A longitudinal systems biology analysis of lactulose withdrawal in hepatic encephalopathy. *Metab. Brain Dis.* **27**, 205–215 (2012).
- Bajaj, J. S. *et al.* Altered profile of human gut microbiome is associated with cirrhosis and its complications. *J. Hepatol.* **60**, 940–947 (2014).
- Bajaj, J. S. *et al.* Modulation of the metabiome by rifaximin in patients with cirrhosis and minimal hepatic encephalopathy. *PLoS ONE* **8**, e60042 (2013).
- Kanno, T. *et al.* Gastric acid reduction leads to an alteration in lower intestinal microflora. *Biochem. Biophys. Res. Commun.* **381**, 666–670 (2009).
- Bajaj, J. S. *et al.* Systems biology analysis of omeprazole therapy in cirrhosis demonstrates significant shifts in gut microbiota composition and function. *Am. J. Physiol. Gastrointest. Liver Physiol.* **307**, G951–G957 (2014).
- Segata, N. *et al.* Metagenomic biomarker discovery and explanation. *Genome Biol.* **12**, R60 (2011).
- Gillevet, P., Sikaroodi, M., Keshavarzian, A. & Mutlu, E. A. Quantitative assessment of the human gut microbiome using multitag pyrosequencing. *Chem. Biodivers.* **7**, 1065–1075 (2010).
- Zhang, Z. *et al.* Large-scale survey of gut microbiota associated with MHE via 16S rRNA-based pyrosequencing. *Am. J. Gastroenterol.* **108**, 1601–1611 (2013).
- Elkrief, L. *et al.* Diabetes mellitus is an independent prognostic factor for major liver-related outcomes in patients with cirrhosis and chronic hepatitis C. *Hepatology* **60**, 823–831 (2014).

**Author Contributions** J.S.B. supervised the patient recruitment, sample collection and clinical analysis. He was also involved in the data interpretation, analysis and drafting of the manuscript. N.S.B. was involved in data analysis and interpretation. P.M.G. was responsible for data analysis, interpretation and drafting of the manuscript. All authors participated in the critical revision of the manuscript.

**Competing Financial Interests** Declared none.

doi:10.1038/nature14851

## Qin *et al.* reply

REPLYING TO J. S. Bajaj, N. S. Betrapally & P. M. Gillevet *Nature* **525**, <http://dx.doi.org/10.1038/nature14851> (2015)

In the accompanying Comment<sup>1</sup>, a concern expressed by Bajaj *et al.* is that diagnostics of liver cirrhosis by microbiome analysis that we report<sup>2</sup> may be mainly due to the microbiome alterations in decompensated patients (DP). To address it we tested how accurately compensated patients (CP) can be diagnosed by microbiome analysis. Two slightly different criteria of identifying these were used, based on absence of ascites and hepatic encephalopathy ( $n = 54$ ) and absence of ascites only ( $n = 57$ ).

First, we constructed a discriminator of patients (P,  $n = 98$ ) and healthy controls (H,  $n = 83$ ) in the discovery cohort, disregarding the patient status (CP or DP). For that we used as input the presence and abundance of 66 metagenomic species (MGS) differentially represented in the two groups<sup>2</sup> and as output area under curve (AUC) of a receiver operator characteristic (ROC) analysis, essentially as described previously<sup>3,4</sup>. The optimal discriminator required 7 MGS only and yielded an AUC of 0.95 for the discovery cohort and of 0.94 for the validation cohort (P  $n = 25$ ; H  $n = 31$ ), values somewhat higher than those observed for the discriminator based on 15 biomarkers<sup>2</sup>. The discriminator stratified the CP ( $n = 54$  or  $n = 57$ ) from H ( $n = 114$ ) as accurately as the DP ( $n = 69$  or  $n = 66$ ), with an AUC of 0.95 for all. This shows that the gut microbiome alterations in the two types of patients have highly similar features. These features are not greatly affected by medication, another concern expressed by Bajaj *et al.*<sup>1</sup>, as the discriminator stratified with a comparable efficiency H ( $n = 114$ ) from P that were taking antiviral medication ( $n = 52$ ) or not ( $n = 71$ ) with an AUC of 0.95 for both; taking  $\beta$ -blockers ( $n = 11$ ) or not ( $n = 112$ ), with an AUC of 0.95 and 0.96, respectively; or taking PPI ( $n = 70$ ) or not ( $n = 53$ ), with an AUC of 0.96 and 0.93, respectively. We suggest that the inability to construct an efficient discriminator of H and CP by

Bajaj *et al.*<sup>1</sup> may be due to an inadequate resolution provided by the broadly used gene encoding the 16S ribosomal RNA, which remains generally at the genus level rather than the species one achieved by quantitative metagenomics we deploy<sup>2–4</sup>.

Notwithstanding the similarity of the gut microbiome alterations in CP and DP, there are also differences between the two groups, as suggested by the association of the disease severity scores and the load of the liver cirrhosis-enriched species<sup>2</sup>. Bajaj *et al.*<sup>1</sup> rightly point out an inaccuracy of the calculation of the model for end-stage liver disease (MELD) score in our report<sup>2</sup>, which refers to previous literature; however, the correction had a modest effect, the statistical significance between the scores of patients with the lowest and the highest LC quartile load being  $P < 2 \times 10^{-5}$  rather than the reported  $P < 1 \times 10^{-5}$ .

To further explore the microbiome alterations in CP and DP we searched for the MGS having a significantly different abundance in the two groups, following the approach used for identifying 66 species enriched in C or P groups<sup>2</sup>. Some 30 such MGS were found in the discovery cohort (CP  $n = 45$ ; DP  $n = 54$ ), but only 13 were not present in the set of 66. All 79 species were used to construct the best discriminator for the discovery cohort. It was based on 14 MGS and stratified the CP and DP of the discovery cohort with an AUC of 0.87 and those of the validation cohort (CP  $n = 9$ ; DP  $n = 16$ ) with an AUC of 0.84.

This analysis confirms our finding that the alterations of the gut microbiome are associated with the severity of the disease. However, it provides no evidence for a saltatory alteration to a different composition upon decompensation, which could confound microbiome analysis, as suggested by Bajaj *et al.*<sup>1</sup>; a gradual alteration with the severity would lead to the same result.

Diabetes has been excluded in the patient enrolment in our study<sup>2</sup>. Furthermore, invasion of the gut by oral species was not observed in



the previous studies of the type-2 diabetes, notwithstanding the use of quantitative metagenomics, which would have easily revealed them were they present<sup>5,6</sup>. Alterations of the gut microbiome owing to liver cirrhosis are therefore unlikely to be confounded by diabetes and the diagnostics of the two pathologies by the gut microbiome analysis remains a real possibility. Short-term nutritional changes, such as hospital diet, generally have only a modest effect on gut microbiota; long term dietary patterns, which affect it more<sup>7,8</sup>, are not very significantly different for the cirrhosis patients and healthy controls in the Chinese population from which the participants enrolled in our study were drawn<sup>2</sup>.

In conclusion, while we adhere to the call of Bajaj *et al.*<sup>1</sup> for caution regarding potential confounders in microbiome analysis, we strongly disagree with their suggestion that the alterations we report are “epiphenomena” rather than actual differences of gut microbial communities associated with liver cirrhosis. We suggest that microbiome analysis might supplant current inadequate clinical diagnostic parameters and/or invasive procedures such as liver biopsy for detecting compensated cirrhosis.

**Nan Qin<sup>1,2\*</sup>, Emmanuelle Le Chatelier<sup>3\*</sup>, Jing Guo<sup>1</sup>, Edi Prifti<sup>3</sup>, Lanjuan Li<sup>1,2</sup> & S. Dusko Ehrlich<sup>3,4</sup>**

<sup>1</sup>State Key Laboratory for Diagnosis and Treatment of Infectious Disease, The First Affiliated Hospital, College of Medicine, Zhejiang University, 310003 Hangzhou, China.

email: ljli@zju.edu.cn

<sup>2</sup>Collaborative Innovation Center for Diagnosis and Treatment of Infectious Diseases, Zhejiang University, 310003 Hangzhou, China.

<sup>3</sup>Metagenopolis, Institut National de la Recherche Agronomique, 78350 Jouy en Josas, France.

email: dusko.ehrlich@jouy.inra.fr

<sup>4</sup>King's College London, Centre for Host-Microbiome Interactions, Dental Institute Central Office, Guy's Hospital, London Bridge, London SE1 9RT, UK.

\*These authors contributed equally to this work.

1. Bajaj, J. S., Betrapally, N. S. & Gillevet, P. M. Decompensated cirrhosis and microbiome interpretation. *Nature* **525**, <http://dx.doi.org/10.1038/nature14851> (2015).
2. Qin, N. *et al.* Alterations of the human gut microbiome in liver cirrhosis. *Nature* **513**, 59–64 (2014).
3. Le Chatellier, E. *et al.* Richness of human gut microbiome correlates with metabolic markers. *Nature* **500**, 541–546 (2013).
4. Cotillard, A. *et al.* Dietary intervention impact on gut microbial gene richness. *Nature* **500**, 585–588 (2013).
5. Qin, J. *et al.* A metagenome-wide association study of gut microbiota in type 2 diabetes. *Nature* **490**, 55–60 (2012).
6. Karlsson, F. H. *et al.* Gut metagenome in European women with normal, impaired and diabetic glucose control. *Nature* **498**, 99–103 (2013).
7. Wu, G. D. *et al.* Linking long-term dietary patterns with gut microbial enterotypes. *Science* **334**, 105–108 (2011).
8. Claesson, M. J. *et al.* Gut microbiota composition correlates with diet and health in the elderly. *Nature* **488**, 178–184 (2012).

doi:10.1038/nature14852

## NEUROSCIENCE

# Forgetfulness illuminated

Memories are stored in the complex network of neurons in the brain. With the help of innovative tools to manipulate the connections between neurons, memories in mice can now be erased with a beam of light. [SEE ARTICLE P.333](#)

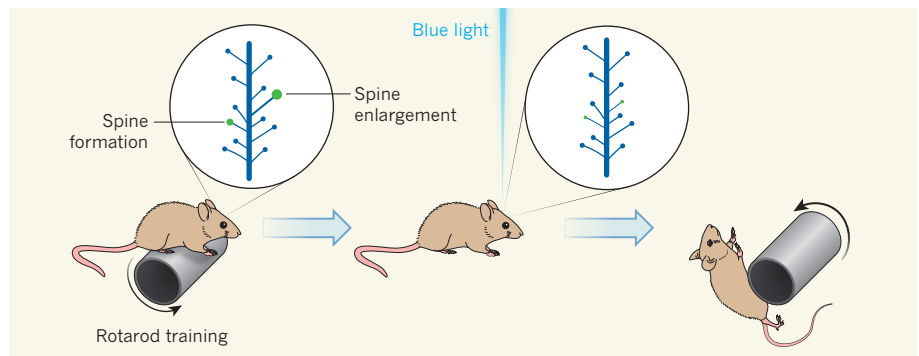
JU LU & YI ZUO

More than a century ago, the German biologist Richard Semon proposed that memories leave physical traces in the brain, and coined the term 'engram' to describe such traces<sup>1</sup>. Although the concept has gained general recognition, the search for the engram is ongoing. In this regard, the synapse — a specialized connecting region between neurons — has received much attention, but there is still no direct evidence of a causal link between synaptic changes and memory formation. In this issue, Hayashi-Takagi *et al.*<sup>2</sup> (page 333) fill this gap. Using ingenious protein engineering and live imaging, the authors identify which synapses are activated when a mouse learns a motor skill, and then weaken these synapses to erase motor memory.

Most synapses in the brain form between axons (neuronal 'output cables') and dendrites (input cables). Signals to excitatory synapses are usually received by micrometre-sized protrusions called spines that emanate from dendrites. The size of the spine head correlates with the strength of the synapse<sup>3</sup>. Spines may emerge, disappear or change in size during learning and memory formation, reflecting changes in the wiring of neuronal circuits<sup>3</sup>.

To investigate the causal relationship between the formation of motor memories and the structural potentiation of spines (spine formation or enlargement), Hayashi-Takagi *et al.* developed an 'optoprobe' called AS-PaRac1 that manipulates potentiated spines in response to light. The DNA construct for AS-PaRac1 encodes a light-activatable version of the small signalling protein Rac1, whose prolonged activity induces spines to shrink. The construct also incorporates the dendrite-targeting sequence of the gene *Arc*, which is expressed rapidly and transiently in response to neuronal activity, ensuring that the probe moves to dendritic spines that are undergoing structural potentiation. The AS-PaRac1 optoprobe is the first optogenetic tool to enable the manipulation of potentiated spines.

Hayashi-Takagi and colleagues expressed AS-PaRac1 in the motor cortex of mice and trained the animals to run on an accelerating rotating rod known as a rotarod. Light



**Figure 1 | Inducing forgetting.** A neuron receives excitatory signals from other neurons through dendritic spines. When a mouse learns a new task, such as running on an accelerating rotating rod (a rotarod), spines involved in learning this task become potentiated (new spines form and existing spines increase in size). Hayashi-Takagi *et al.*<sup>2</sup> developed an 'optogenetic construct' based on a light-activatable form of the small signalling protein Rac1, which targets recently potentiated dendritic spines. Blue light activates the modified Rac1, which induces shrinkage of the spines. The authors found that spine shrinkage caused the mouse to forget the skill it had learnt, so it soon fell off the rotating rod.

activation of AS-PaRac1 in potentiated spines after learning caused the spines to shrink, disrupting the animals' ability to run on the rotarod. This demonstrates the causal relationship between synaptic strength and motor memory in this context (Fig. 1).

Next, the authors showed that the effect of the probe is task-specific. When mice learnt to run on the rotarod and then learnt to walk on a thin beam, disrupting the spines that were potentiated during beam walking did not affect performance on the rotarod. Furthermore, AS-PaRac1 activation in spines that spontaneously potentiated two days after learning (presumably because of unrelated motor tasks) did not affect motor performance. Finally, when the authors retrained mice on the same task for which spine potentiation had been disrupted, most of the optically shrunken spines reverted to their original potentiated sizes. Together, these results suggest that distinct subsets of synapses are altered in a task-specific way during motor learning and memory formation.

In the long quest for the engram, neuroscientists have reached the consensus that the mammalian brain stores different memory traces in different subsets of neurons in specific regions. Methods for labelling, imaging, activating and silencing neurons in animals have enabled researchers to map the ensemble

of neurons that correlates with a particular learning task, to manipulate their activities, and even to generate artificial memory traces<sup>4–6</sup>. However, a single neuron may participate in the processing and storage of more than one distinct piece of information<sup>7</sup>. Therefore, the engram of a particular memory involves not only the identity of the constituent neurons, but also the entire set of synaptic connections between these neurons. How memory is allocated at this synaptic level remains unclear.

To qualify as an engram, a synaptic circuit should satisfy several criteria. First, changes in synaptic structures and function should correlate with learning. Second, blocking such synaptic modifications should prevent memory formation, demonstrating the need for these changes. And third, artificially inducing synaptic changes should be sufficient to produce a memory without the need for behavioural training. Over the past decade, *in vivo* imaging has revealed<sup>8</sup> that the dynamic formation and elimination of dendritic spines correlates with motor-skill learning and memory. Now, Hayashi-Takagi and colleagues have taken the next step, by establishing necessity — they show that undoing the synaptic changes that accompany motor learning does indeed disrupt the memory.

The development of genetic and optical tools such as AS-PaRac1 promises to enable



dissection of the finer details of the engram. The use of promoter sequences that drive the expression of target genes in a cell-type-specific manner, as well as connectivity-specific labelling methods<sup>9</sup>, can help to unravel the roles in learning and memory of synaptic circuits formed by different types of neuron — revealing, for example, the relative contributions of excitatory and inhibitory neurons, or of neurons in different layers of the brain's cortex. When we have a deeper understanding of the molecular signalling events that occur at synapses during memory formation<sup>10</sup>, tools similar to AS-PaRaC1 can be devised to modulate other components of the molecular machinery. Improved microscopy techniques can already target individual neurons or synapses<sup>11</sup>, rather than manipulating a population of neurons as a whole.

When used together, such technical advances will enable us to strengthen existing engrams, to facilitate the formation of new ones, and to generate synthetic memory traces at the synaptic level. We will then be able to study the interaction between different memory traces, as well as the mechanisms that translate an engram into behavioural outputs. These efforts should allow us to gain an understanding of the intriguing phenomenon of memory simply by shining a light on its physical basis. ■

Ju Lu and Yi Zuo are in the Department of Molecular, Cell and Developmental Biology, University of California, Santa Cruz, Santa Cruz, California 95064, USA. e-mails: jlu39@ucsc.edu; yizuo@ucsc.edu

## CATALYSIS

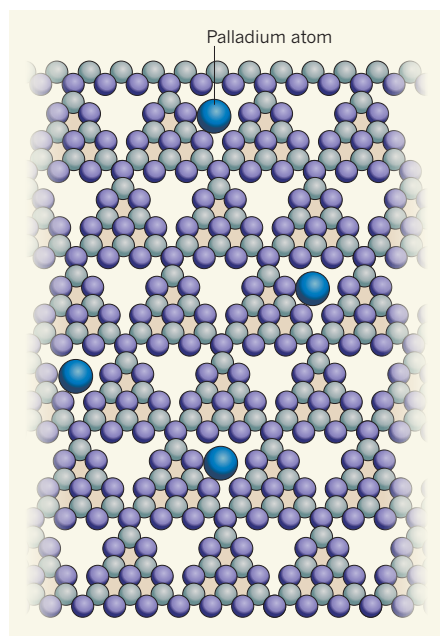
# Tens of thousands of atoms replaced by one

Many catalysts comprise metal nanoparticles on solid supports. The discovery that single atoms of palladium anchored to a solid support also exhibit high catalytic activity might help to conserve the supply of this and related rare metals.

JOHN MEURIG THOMAS

The platinum-group metals — ruthenium, rhodium, palladium, osmium, iridium and platinum — are extensively used as catalysts in industries that produce compounds such as agrochemicals, dyestuffs and pharmaceuticals, and several of them are crucial components of catalytic converters in cars. But as demand for these relatively scarce metals increases, their future availability is a cause for concern. This would be dispelled if the metals could be used in an atomically dispersed state, rather than as nanoparticles containing up to 100,000 atoms, as is conventional. Writing in *Angewandte Chemie*, Vilé *et al.*<sup>1</sup> report that individual atoms of palladium can be anchored to carbon nitride ( $C_3N_4$ ), an easily prepared nanoporous solid<sup>2</sup>. The resulting materials are excellent, thermally stable catalysts for selective hydrogenation reactions, which facilitate the production of many organic substances, including polymers and biologically important compounds<sup>3</sup>.

There are many examples of catalysts in which the active components are supported nanoparticles of platinum-group metals (PGMs) or gold (see refs 4–6, for example). But in several cases, it has long been suspected<sup>7–9</sup> that the nanoparticles are unimportant, and that catalysis occurs at single-atom sites. Indeed, isolated metal atoms have previously



**Figure 1 | A single-atom palladium catalyst.** Vilé *et al.*<sup>1</sup> report that isolated palladium atoms on a solid support of carbon nitride ( $C_3N_4$ ; carbon atoms, grey; nitrogen atoms, purple) act as catalysts for hydrogenation reactions. Strong bonds to the nitrogen atoms firmly anchor the palladium atoms in roughly triangular pores in the stacked, two-dimensional layers of the support. Only one layer is depicted, for simplicity. (Adapted from ref. 1.)

1. Semon, R. *Die Mneme als erhaltendes Prinzip im Wechsel des organischen Geschehens* (Wilhelm Engelmann, 1904)
2. Hayashi-Takagi, A. *et al. Nature* **525**, 333–338 (2015).
3. Holtmaat, A. & Svoboda, K. *Nature Rev. Neurosci.* **10**, 647–658 (2009).
4. Han, J.-H. *et al. Science* **323**, 1492–1496 (2009).
5. Garner, A. R. *et al. Science* **335**, 1513–1516 (2012).
6. Ramirez, S. *et al. Science* **341**, 387–391 (2013).
7. Jia, H., Rochefort, N. L., Chen, X. & Konnerth, A. *Nature* **464**, 1307–1312 (2010).
8. Chen, C.-C., Lu, J. & Zuo, Y. *Front. Neuroanat.* **8**, 28 (2014).
9. Luo, L., Callaway, E. M. & Svoboda, K. *Neuron* **57**, 634–660 (2008).
10. Mayford, M., Siegelbaum, S. A. & Kandel, E. R. *Cold Spring Harb. Perspect. Biol.* **4**, a005751 (2012).
11. Packer, A. M., Russell, L. E., Dalgleish, H. W. P. & Häusser, M. *Nature Methods* **12**, 140–146 (2015).

This article was published online on 9 September 2015.

been manipulated as a strategy for enabling selective catalytic hydrogenations<sup>10</sup>: atomically dispersed palladium atoms on the surfaces of a copper crystal stimulate local breaking of the bonds in hydrogen molecules, and the resulting hydrogen atoms become mobile on the copper surface, readily reacting with unsaturated molecules such as acetylene and styrene. However, in that system, the single atoms are laid down on the copper by heating a palladium source in a high-vacuum chamber using an electron beam. This method is suitable for preparing single-atom catalysts of other PGMs, but does not readily translate to the production of industrial-scale quantities of catalysts.

In their study, Vilé and colleagues propose that the catalytically active individual palladium atoms are tenaciously attached to the nitrogen atoms of the  $C_3N_4$  support (Fig. 1), owing to the lone pair of electrons that each nitrogen atom has<sup>11</sup>. The authors' X-ray-absorption studies found no evidence of palladium–palladium bonds, indicating that the atoms are indeed separate from each other. The researchers also studied their samples using a technique called annular dark-field electron microscopy<sup>12</sup>, which takes advantage of the Rutherford scattering of electrons<sup>13</sup> (scattering at large angles) to detect heavy atoms of PGMs on the light elements of  $C_3N_4$ . These experiments identified only single palladium atoms in the active catalyst.

Vilé and co-workers' catalysts are particularly notable because reproducible, thermally stable single-atom preparations can be readily made, provided that care is taken to incorporate only small amounts of the palladium on the nanoporous support. Moreover,  $C_3N_4$  is inexpensive and may be routinely prepared in a graphite-like form<sup>2,14</sup> that has relatively widely separated layers, thereby increasing the accessibility of the anchored palladium atoms to reactants. The authors report that it also has the merit of a high surface area

(about 150 square metres per gram), which maximizes catalytic performance.

The main hydrogenation reaction studied by the authors was the conversion of 1-hexyne to 1-hexene, in which carbon–carbon triple bonds are selectively converted to double bonds, but not further to single ones. Such selective hydrogenations have conventionally used a Lindlar catalyst<sup>15</sup>, which consists of nanoparticles of a palladium–lead compound<sup>16</sup> on a calcium carbonate support. The authors' single-atom palladium catalyst enables much higher yields and faster reactions than either a Lindlar catalyst or hydrogenation catalysts based on nanoparticles of platinum or gold<sup>17</sup>. It also yields 1-hexene with greater than 99% selectivity. Moreover, after repeated use (five successive tests), the catalyst displays no decrease in selectivity nor in the fraction of 1-hexyne converted to 1-hexene. Finally, Vilé and colleagues report that their single-atom catalyst enables the hydrogenation of nitrobenzene to form aniline exclusively. This kind of reaction is used to make compounds for the dyestuffs industry and key intermediates in the manufacture of agrochemicals and pharmaceuticals.

Single-atom solid catalysts are of considerable interest, from both a practical<sup>9,18,19</sup> and a theoretical<sup>20</sup> perspective, not only because selective hydrogenations are among the most valuable conversions in industrial chemistry, but also because these reactions are atom efficient<sup>7</sup>: the minimum number of atoms is used in each reaction, reducing waste. Such catalysts can also be used for other reactions. For example, recent work<sup>18</sup> shows that single atoms of platinum function as atom-efficient catalysts for the water–gas shift reaction, which is used to generate pure hydrogen for the synthesis of ammonia. Single-atom platinum catalysts have also been used for the selective hydrogenation of nitroaromatic compounds<sup>9</sup>.

Readily prepared single-atom platinum catalysts<sup>9</sup> supported on iron oxide (FeO<sub>x</sub>) have been reported to be much more active, selective and durable than analogous nanoparticle platinum catalysts. Remarkably, single atoms of platinum supported on FeO<sub>x</sub> are chemoselective in the hydrogenation reactions that they catalyse — that is, they can discriminate between two or more regions of a molecule that could potentially react with hydrogen. For example, the catalysts convert nitro groups (NO<sub>2</sub>) to amino groups (NH<sub>2</sub>), but leave carbonyl groups (C=O) and benzene rings untouched. In one such reaction, a single-atom platinum catalyst displayed a turnover frequency (the number of reactant molecules converted to product per unit of time) of 1,500 per hour, which is 20 times as high as the previous best result reported in the literature. The selectivity for the substrate of that reaction was about 99%, the highest reported for any PGM catalyst.

The future looks bright for the use of PGMs

as catalysts, both on laboratory and industrial scales, because the preparation of most kinds of single-atom metal catalyst is likely to be straightforward, and because characterization of such catalysts has become easier with the advent of techniques that readily discriminate single atoms from small clusters and nanoparticles. A prerequisite is to find ways of securely anchoring single atoms of these expensive metals to high-surface-area, cheap and plentiful solids composed of elements that are abundantly available, as Vilé *et al.* and others<sup>18</sup> have done. If this can be achieved generally, then the future deployment of PGMs in solid catalysts will be transformed. ■

**John Meurig Thomas** is in the Department of Materials Science and Metallurgy, and at Peterhouse, University of Cambridge, Cambridge CB3 0FS, UK.  
e-mail: jmt2@cam.ac.uk

- Vilé, G. *et al.* *Angew. Chem. Int. Edn* **54**, 11265–11269 (2015).
- Goettmann, F., Fischer, A., Antonietti, M. & Thomas, A.

- Angew. Chem. Int. Edn* **45**, 4467–4471 (2006).
- Thomas, J. M., Johnson, B. F. G., Raja, R., Sankar, G. & Midgley, P. A. *Acc. Chem. Res.* **36**, 20–30 (2003).
- Valdez, M., Lai, X. & Goodman, D. W. *Science* **281**, 1647–1650 (1998).
- Hughes, M. D. *et al.* *Nature* **437**, 1132–1135 (2005).
- Haruta, M. *Faraday Disc.* **152**, 11–32 (2011).
- Thomas, J. M. *Design and Applications of Single Site Heterogeneous Catalysts* (Imperial Coll. Press, 2012).
- Fu, Q., Saltsburg, H. & Flytzani-Stephanopoulos, M. *Science* **301**, 935–938 (2003).
- Wei, H. *et al.* *Nature Commun.* **5**, 5634 (2014).
- Kyriakou, G. *et al.* *Science* **335**, 1209–1212 (2012).
- Arrigo, R. *et al.* *ACS Catal.* **5**, 2740–2753 (2015).
- Krivanek, O. L. *et al.* *Nature* **464**, 571–574 (2010).
- Midgley, P. A., Weyland, M., Thomas, J. M. & Johnson, B. F. G. *Chem. Commun.* 907–908 (2001).
- Groenewolt, M. & Antonietti, M. *Adv. Mater.* **17**, 1789–1792 (2005).
- Lindlar, H. *Helv. Chim. Acta* **35**, 446–450 (1952).
- Palczewska, W., Jablonski, A., Kaszkur, Z., Zuba, G. & Wernisch, J. J. *Mol. Catal.* **25**, 307–316 (1984).
- Serna, P., Boronat, M. & Corma, A. *Top. Catal.* **54**, 439–446 (2011).
- Yang, M. *et al.* *J. Am. Chem. Soc.* **137**, 3470–3473 (2015).
- Thomas, J. M. *Phil. Trans. R. Soc. A* (in the press).
- Peters, B. & Scott, S. L. *J. Chem. Phys.* **142**, 104708 (2015).

#### EVOLUTIONARY BIOLOGY

## Perplexing effects of phenotypic plasticity

**Research on guppies provides evidence that phenotypic plasticity — an organism's ability to alter its characteristics in response to changes in the environment — can both constrain and facilitate adaptive evolution. SEE LETTER P.372**

JUHA MERILÄ

**A**ltered or new environmental conditions, such as those brought about by climate change, are important sources of selection pressures that drive organismal adaptation and evolution. But alongside genetic adaptation, organisms can respond to environmental challenges through adaptive phenotypic plasticity, which refers to a non-genetic shift in the average characteristics (phenotype) of a population towards an evolutionary optimum. Whether phenotypic plasticity generally facilitates or constrains adaptive (genetic) evolution remains a contentious issue<sup>1–4</sup>. On page 372 of this issue, Ghalambor *et al.*<sup>5</sup> provide experimental evidence from guppies suggesting that adaptive phenotypic plasticity in gene-expression patterns constrains evolution. But they also find that non-adaptive plasticity — phenotypic changes that do not directly contribute to increased fitness under the changed conditions — may facilitate adaptive genetic change by increasing the strength of natural selection.

The authors' experiments involved transplanting wild Trinidadian guppies (*Poecilia reticulata*; Fig. 1) from a stream that also hosted predatory cichlid fish into two replicate streams without cichlids. They then compared patterns of brain gene expression between the introduced and original (ancestral) populations after three or four generations. Parallel changes in gene expression had occurred for 135 genes in the two introduced populations, and these new levels of gene expression were similar to those exhibited by a native cichlid-free population. This suggested rapid adaptive evolution in the introduced populations.

However, the evolved differences were mostly (89% of the genes) in the opposite direction to that of phenotypic plasticity in expression patterns in the ancestral population. This was inferred by comparing the gene expression in ancestral fish reared in either the presence or absence of chemical cues from predatory cichlids. Thus, the phenotypic plasticity in these genes can be considered non-adaptive. The remaining 11% of genes exhibited adaptive plasticity — the evolved differences in gene expression in the





**Figure 1 | Trinidadian guppies.** Ghalambor *et al.*<sup>5</sup> show that high levels of non-adaptive phenotypic plasticity in a source population seem to facilitate rapid adaptive evolution of gene-expression patterns in guppies transplanted into a different environment.

experimentally introduced populations were concordant with the direction of change of expression levels in ancestral fish raised in the absence of predatory-fish cues. The authors also observed that there was little or no population divergence in the expression of these genes in either of the introduced populations.

The latter findings support evolutionary models predicting that adaptive phenotypic plasticity should weaken the strength of directional selection and thereby slow the rate of evolution (see refs 6 and 7 for examples). However, the real stunner of the study was the discovery that most of the evolved (genetic) differences in gene-expression patterns in the introduced guppy populations had taken place in the opposite direction to the direction of plasticity in the ancestral population. This inverse relationship between the direction of plasticity and the direction of adaptive evolution suggests that non-adaptive plasticity may facilitate (in the authors' words, potentiate) evolution by increasing the strength of directional selection required to create the observed divergence in gene-expression patterns.

The authors obtained support for the hypothesized increase in directional selection against non-adaptive plasticity by examining evolutionary changes in the magnitude of plasticity (quantified as the mean difference in expression levels of gene transcripts in the predator-cue-treated groups) between ancestral and introduced populations. Although phenotypic plasticity is, by definition, a

non-genetic response to environmental cues, the capacity to express it, and its magnitude, can be genetically variable<sup>1,4</sup>. Consequently, if directional selection had acted most strongly on gene transcripts exhibiting non-adaptive plasticity, then the magnitude of plasticity in introduced populations in response to this selection should be reduced. This was just what Ghalambor *et al.* observed. Moreover, the decline in the magnitude of plasticity in the introduced populations was inversely proportional to plasticity in the ancestral population. This also aligns with the expectation that transcripts exhibiting the greatest non-adaptive plasticity should be the ones that are most strongly selected against.

Although the findings that phenotypic plasticity can both constrain and facilitate evolutionary (genetic) adaptation are not unprecedented, several features of Ghalambor and colleagues' study set it apart from earlier work on this topic. For instance, instead of focusing on a limited number of traits, the authors assessed the plasticity of a large number of traits (expressed genes), which allowed them to draw robust quantitative conclusions. Nevertheless, a question to be addressed is whether results from gene-expression analyses can be extended and generalized to macroscopic traits that have more-direct ecological relevance. Similarly, most previous empirical studies that focused on the direction of plastic responses and the direction of subsequent evolutionary divergence in wild populations have been limited to comparisons between

ancestral and derived populations long after they diverged. The new study's focus on initial patterns of plasticity and subsequent rapid adaptive divergence in the wild provides a thought-provoking complement to laboratory experiments that have provided evidence supporting both positive (adaptive)<sup>8,9</sup> and negative (non-adaptive)<sup>10</sup> relationships between the directions of plastic responses and evolution.

Ghalambor and colleagues' results are also intriguing because most (but not all) attempts to model the effects of plasticity on subsequent evolution have assumed it to be adaptive. Thus, the observed negative relationship between the direction of plasticity and the direction of evolution in guppies may guide future theoretical work in the field. Furthermore, although increased strength of selection caused by non-adaptive plasticity may contribute to rapid adaptation and increase the likelihood of population persistence, it may also lead to reduced population size and an increased risk of demographic collapse<sup>3</sup>. By reducing population size, selection stemming from non-adaptive plasticity may expose a population to an increased rate of random genetic changes owing to a process known as genetic drift. This would in turn propagate loss of genetic variation and reduced efficiency of selection, counteracting the proposed benefit from non-adaptive plasticity.

As fascinating as it is to suggest that maladaptive plasticity may be a strong driver of evolution, sceptics may require further experimental studies from the wild with more population replicates and with a focus on traits with established ecological relevance (such as behaviours and morphology) to be convinced. Such studies would also be helpful, if not essential, in developing parameters for models that aim to understand how the interplay between phenotypic plasticity, natural selection and random genetic drift influences evolutionary changes. ■

**Juha Merilä** is in the Department of Biosciences, University of Helsinki, 00014 Helsinki, Finland.  
e-mail: juha.merila@helsinki.fi

1. West-Eberhard, M. J. *Developmental Plasticity and Evolution* (Oxford Univ. Press, 2003).
2. Price, T. D., Qvarnström, A. & Irwin, D. E. *Proc. R. Soc. B* **270**, 1433–1440 (2003).
3. Chevin, L.-M., Lande, R. & Mace, G. M. *PLoS Biol.* **8**, e1000357 (2010).
4. Pfennig, D. W. *et al. Trends Ecol. Evol.* **25**, 459–467 (2010).
5. Ghalambor, C. K. *et al. Nature* **525**, 372–375 (2015).
6. Ancel, L. W. *Theor. Popul. Biol.* **58**, 307–319 (2000).
7. Paenke, I., Sendhoff, B. & Kawecki, T. J. *Am. Nat.* **170**, E47–E58 (2007).
8. Waddington, C. H. *Adv. Genet.* **10**, 257–293 (1961).
9. Suzuki, Y. & Nijhout, H. F. *Science* **311**, 650–652 (2006).
10. Schaum, C. E. & Collins, S. *Proc. R. Soc. B* **281**, 20141486 (2014).

This article was published online on 2 September 2015.

## CANCER

# Repositioned to kill stem cells

**Chemotherapy-resistant cancer stem cells make it hard to cure many forms of the disease. Repositioning an existing drug to tackle this problem could significantly improve treatment for one form of leukaemia. [SEE LETTER P.380](#)**

TESSA HOLYOAKE & DAVID VETRIE

In most cases of chronic myeloid leukaemia (CML), a daily oral medication can rapidly transform a progressive and ultimately fatal cancer into a chronic but manageable condition. But this is not a cure. The persistence of quiescent (dormant, non-cycling) and thus drug-resistant leukaemic stem cells (LSCs) poses an unmet clinical challenge, and any attempt to cure CML must target the eradication of these cells. In this issue, Prost *et al.*<sup>1</sup> (page 380) present provocative preclinical and early clinical findings demonstrating that a drug currently used for diabetes therapy can be repositioned to target a pathway that controls quiescence in LSCs, causing the gradual erosion of this cellular pool.

The cause of CML is a mutation in a normal blood stem cell involving an exchange

of genetic material between chromosomes 9 and 22. This translocation creates a cancer-driving gene known as *BCR-ABL1*, which produces a protein with enhanced activity as a tyrosine kinase enzyme, leading to uncontrolled cell proliferation. *BCR-ABL1* has been shown to be sufficient to drive the development of leukaemia in mouse models<sup>2</sup>, and the discovery of this protein led to the development of tyrosine kinase inhibitors (TKIs) for CML treatment.

In the past two decades, TKIs have dramatically improved the outcome for people with this cancer. Most of those who present with early disease respond rapidly to TKI therapy and go into long-lasting remission. However, TKIs fail to eradicate LSCs, the cells that initiate and maintain CML, and these drug-resistant cells can drive relapse, or evolve to cause further forms of TKI resistance and

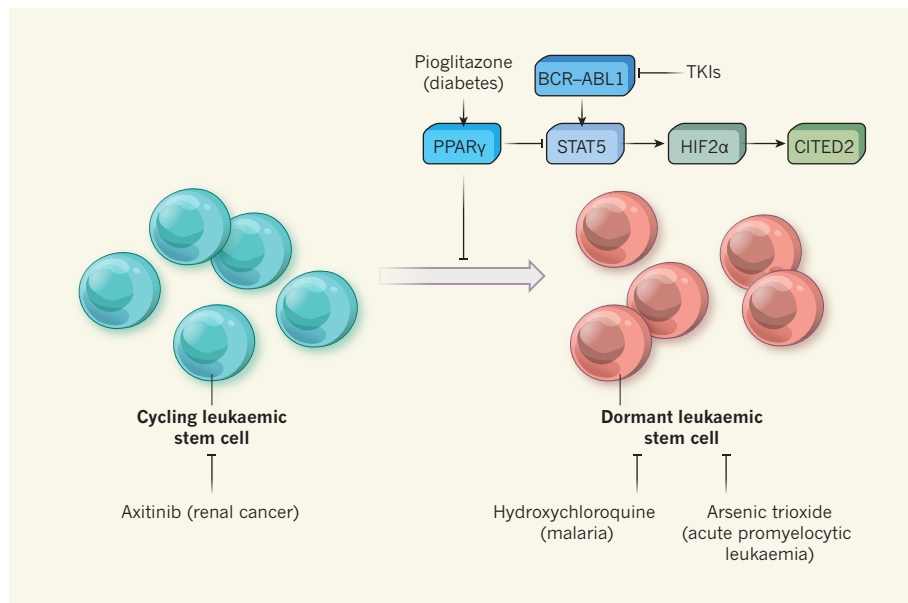
more-aggressive disease. As a result, people on life-long TKI therapy are exposed to associated, often serious, side effects and may cease to respond to the treatment at any time. Furthermore, the significantly improved survival for those taking TKIs means that the prevalence of CML is increasing each year, with inherent social and economic implications.

Several potential mechanisms to explain the insensitivity of LSCs to TKIs have been proposed, including cellular quiescence. Prost *et al.* report that quiescence in LSCs is regulated by a pathway involving the receptor PPAR $\gamma$ , the transcription factors STAT5 and HIF2 $\alpha$ , and the protein CITED2, which is known<sup>3</sup> to regulate blood stem-cell quiescence (Fig. 1). A particular strength of the study was the use of primary blood stem cells (expressing the marker CD34) from people with CML to dissect the pathway and confirm the role of each component in regulating LSC quiescence.

The authors go on to show that combining imatinib, the standard TKI used to manage CML, with the antidiabetic agent pioglitazone, which activates PPAR $\gamma$ , blocks this pathway in CML cells. The synergistic effects of the drugs reduce STAT5 expression and activity, down-regulate HIF2 $\alpha$  and CITED2 expression, and trigger the death of quiescent LSCs. Although the mechanism by which LSCs are killed in response to this drug combination is not clear, they are probably either killed directly or driven to exit quiescence, which may lead to their eradication by the TKI. The authors also demonstrate that the compound JQ1, a bromodomain inhibitor with broad activity that includes the suppression of STAT5 activity, is as effective as pioglitazone (in combination with imatinib). Although this finding supports a role for the STAT5 pathway in LSC quiescence, the door is still open for studies of other agents that may target LSCs through this or alternative pathways.

Collectively, these results strengthen the concept that cancer stem cells exhibit vulnerabilities in otherwise normal molecular pathways that may be targeted in a selective manner to obtain a cure. Earlier work demonstrated that CML stem-cell quiescence is in part maintained by the promyelocytic leukaemia tumour-suppressor protein, which can be targeted by arsenic trioxide<sup>4</sup>, and that the cellular process of autophagy functions as a survival pathway for CML stem cells that can be targeted by repositioning the anti-malarial agent hydroxychloroquine<sup>5</sup> (Fig. 1). Both of these approaches are currently under investigation in the clinic.

Prost *et al.* also tested the addition of pioglitazone to imatinib therapy in three people with CML, and found that they converted from having demonstrable residual leukaemia to being disease-free. The effect lasted for months to years after pioglitazone treatment



**Figure 1 | Targeting leukaemic stem cells in chronic myeloid leukaemia.** Prost *et al.*<sup>1</sup> describe a molecular pathway, involving the receptor PPAR $\gamma$ , the transcription factors STAT5 and HIF2 $\alpha$ , and the regulatory protein CITED2, that induces leukaemic stem cells (LSCs) to enter a dormant (quiescent) state. They also show that the drug pioglitazone, approved for diabetes treatment, activates PPAR $\gamma$  to block this pathway, and can kill these cells when used in conjunction with tyrosine kinase inhibitors (TKIs), which inhibit the protein BCR-ABL1 and thus STAT5, and which are the standard therapy against active (cycling) leukaemic cells. Several drugs used to treat other diseases, such as axitinib, arsenic trioxide and hydroxychloroquine, have also been repositioned to treat chronic myeloid leukaemia, but these have different mechanisms of action.



ceased. These data provided a strong rationale for a phase II clinical trial, which started in July 2009 (ACTIM EudraCT 2009-011675-79). Although the interim results from this trial are encouraging, the study is non-randomized, so it will be difficult to ascertain definitively that improved response rates are driven by pioglitazone.

Despite the need for further clinical testing of this combination therapy, Prost *et al.* have demonstrated the substantial potential for drug repositioning in CML research. Their results follow a recent report<sup>6</sup> in which axitinib, a TKI approved for the treatment of drug-resistant renal-cell cancer, was repositioned to tackle TKI resistance in CML. Using drugs that have already been approved for other purposes can shorten the drug-development pathway by 5–10 years and reduce risks and costs.

Although drug repositioning can be rather serendipitous, Prost and colleagues had a tangible rationale that PPAR $\gamma$  activators such as pioglitazone warranted investigation in CML on the basis of their observation<sup>7</sup> of the drugs' activity against a cell-line model of the disease. Already around 30% of drugs newly approved for a particular treatment have been repositioned from another therapy, and such hypothesis-driven repositioning strategies are likely to become more common in cancer drug discovery. This figure is set to rise further as our understanding of cellular pathways and processes increases and we include innovative computational approaches to facilitate disease-, drug- and treatment-oriented drug repositioning. It is clear that repositioning will increasingly help the fast-tracking of drugs into the clinic. As demonstrated by

Prost and colleagues, this could soon signal the beginning of the end for stem-cell quiescence in CML and other cancers. ■

**Tessa Holyoake and David Vetrie** are at the Institute of Cancer Sciences, University of Glasgow, Glasgow G12 0ZD (T.H.) and Glasgow G61 1QH (D.V.), UK. e-mails: tessa.holyoake@glasgow.ac.uk; david.vetrie@glasgow.ac.uk

1. Prost, S. *et al.* *Nature* **525**, 380–383 (2015).
2. Daley, G. Q., Van Etten, R. A. & Baltimore, D. *Science* **247**, 824–830 (1990).
3. Kranc, K. R. *et al.* *Cell Stem Cell* **5**, 659–665 (2009).
4. Ito, K. *et al.* *Nature* **453**, 1072–1078 (2008).
5. Bellodi, C. *et al.* *J. Clin. Invest.* **119**, 1109–1123 (2009).
6. Pemovska, T. *et al.* *Nature* **519**, 102–105 (2015).
7. Prost, S. *et al.* *J. Clin. Invest.* **118**, 1765–1775 (2008).

This article was published online on 2 September 2015.

## CONDENSED-MATTER PHYSICS

# Charge topology in superconductors

**X-ray images of cuprate superconductors reveal the fractured, defect-riddled backbone on which superconductivity develops. The results take us a step closer to understanding how supercurrent flows on small spatial scales. SEE LETTER P.359**

ERICA W. CARLSON

The quantum motion of electrons enforces a high degree of homogeneity in conventional materials such as metals and semiconductors. As a result, the electrons spread out evenly in these materials, like liquid filling a container. By contrast, nanoscale images of copper oxide (cuprate) superconductors have revealed that the materials' electrons form clumps at the surface<sup>1</sup>. On page 359 of this issue, Campi *et al.*<sup>2</sup> report X-ray images of a cuprate superconductor, revealing complex patterns of electrons<sup>3</sup> that are also scaffolded throughout the interior of the material, and on a much larger scale than has been observed before. Just like the skeleton of a coral reef, where, the greater the scale on which the reef is observed, the more complexity meets the eye, the electrons in these materials form structures full of gnarled hollows of varying size.

Campi and colleagues find that the size of the patterns formed by electrons is strongly tied to the degree to which the cuprate superconductor is doped — meaning that a small amount of one type of atom is substituted for another, to change the charge available for conducting current through the material. Materials that have a uniform distribution of electrons, such as semiconductors, are typically robust against spatial variations in

doping level. Nanostructures are a notable exception: variations in doping level affect the performance of the smallest semiconductor devices. The electrons inside cuprate compounds that superconduct at high temperatures (up to 160 kelvin) spontaneously form nanostructures, and so these materials are sensitive to local doping variations.

The authors made their discovery using a technique called scanning micro X-ray diffraction. In this approach, as high-energy light moves through a material, it bends at an angle that depends on the periodicity of local charge variations in the material. It was already known that the charge in cuprate superconductors is often locally ordered into a unidirectional pattern, with a periodicity of about four crystalline unit cells (the smallest periodically repeating structures in a crystal), so that the electron-density distribution resembles striped wallpaper<sup>1,4,5</sup>. Campi *et al.* scan a micrometre-sized X-ray beam across a superconducting sample to probe how the character of this stripy charge-density wave varies from spot to spot in the

**The disordered states that the authors discovered might be exploited to manipulate superconductivity.**

material. But instead of a single sheet of 'striped wallpaper', they find rips, tears and patches in the electronic texture, as though someone had papered a wall with sheets of many different sizes and shapes, and with complete disregard for whether the borders matched up.

As in nanostructure semiconductors, these electronic textures in cuprate superconductors are sensitive to local variations in the doping level on nanometre and micrometre scales. For example, the lower the level of local oxygen doping, the higher is the contrast of the charge-density waves (bright wallpaper stripes), introducing greater disorder into the overall charge pattern.

These effects are important for superconductivity because of two factors: dimensionality and connectivity. The behaviour of electrons ultimately depends on the shape of their quantum-mechanical waves, and waves show vastly different behaviour in different dimensions. In three dimensions, the energy carried by a wave over a distance  $r$  spreads out as  $r^{-2}$ , as in sound waves coming from a speaker. In two dimensions, such as in ripples emanating from a pebble thrown into a pond, the distance dependence changes to  $r^{-1}$ . In one dimension, waves cannot dissipate by spreading out. Like the bow waves of canal tugboats, there is only one way for a wave to go in one dimension: forward. The charge-density structures that Campi *et al.* find are enticing, because they effectively reduce the dimensionality that electrons can explore, which can lead to mechanisms of superconductivity that are fundamentally different from those of conventional superconductors<sup>6,7</sup>.

However, the dimensionality that Campi *et al.* infer is not integer. They find that the sizes of the patches formed by the charge-density waves are distributed according to a power law, which is typical of fractal dimensions. Like any good fractal, these patterns display similarities whether they are observed from close up or far away. Although much

theoretical effort has gone into understanding electrons in three, two and one dimensions, we know little about the behaviour of electrons in fractal dimensions.

The other key ingredient of the effects that the authors observe is connectivity. Macroscopic superconductivity is ultimately a charge-transport phenomenon (for electricity to flow, electrons must be transported from one side of the sample to the other), and this transport is dominated by connectivity. Without connections between different domains, supercurrent cannot flow through the sample, and the material fails to be a practical, bulk superconductor.

A disordered charge distribution can therefore be devastating to connectivity in superconducting materials. One crucial connection erased by disorder can unlink an entire system. Disorder can also affect the nature of the changes in physical properties that accompany the onset of phase transitions (including superconductivity), by smearing out an abrupt

transition, lowering the temperature at which it happens or changing the geometry of the fractal charge distribution associated with a smooth phase transition. Such disorder can make it much harder for a system to equilibrate, causing the changes in its properties to lag in response to external inputs (hysteresis) and to be dependent on past inputs (memory effects). However, these effects can also be turned into an opportunity to control domain morphology through system-training protocols<sup>8</sup>, much like the way in which commercial permanent magnets are prepared in a magnetized state. This means that the disordered states that the authors have discovered might be exploited to manipulate superconductivity along similar lines.

One limitation of Campi and colleagues' study is that they did not directly observe the morphology of the path that the superconducting electrons take. Rather, they inferred it from the morphology of the observed variable charge distribution. More data are needed to

probe that intermediate length between the nanometre and macroscopic length scales, so as to chart the true path of the superconducting electrons. Future work should also investigate how the spatial pathways of superconductivity are affected by the complex interplay between disordered electron distributions and charge-density waves. ■

**Erica W. Carlson** is in the Department of Physics and Astronomy, Purdue University, West Lafayette, Indiana 47907-2036, USA. e-mail: ewcarlson@purdue.edu

1. Kohsaka, Y. *et al.* *Science* **315**, 1380–1385 (2007).
2. Campi, G. *et al.* *Nature* **525**, 359–362 (2015).
3. Dagotto, E. *Science* **309**, 257–262 (2005).
4. Tranquada, J. M., Sternlieb, B. J., Axe, J. D., Nakamura, Y. & Uchida, S. *Nature* **375**, 561–563 (1995).
5. Comin, R. *et al.* *Science* **343**, 390–392 (2014).
6. Emery, V. J., Kivelson, S. A. & Zachar, O. *Phys. Rev. B* **56**, 6120 (1997).
7. Senthil, T. & Fisher, M. P. A. *Phys. Rev. Lett.* **86**, 292–295 (2001).
8. Carlson, E. W. & Dahmen, K. A. *Nature Commun.* **2**, 379 (2011).

cardiovascular disease, the IER curve is steeper at low exposures (implying that the mortality effects of increases in PM<sub>2.5</sub> are greater at lower particulate levels), but generally flattens at higher exposures. Large uncertainties in the IER for PM<sub>2.5</sub> occur in the exposure range of approximately 30–100 micrograms per cubic metre (ref. 7), because no information for cardiovascular mortality due to outdoor PM<sub>2.5</sub> is available, and because only a few studies of second-hand smoke exposure exist. A caveat to Lelieveld and colleagues' estimates of premature deaths from commercial and residential energy sources in Asian countries is that they fall mostly in these areas of high uncertainty.

Studies of the effects of biomass burning on cardiovascular disease or stroke at any level of exposure are also lacking<sup>8</sup>. Furthermore, the largest study so far to examine how sources of fine-particle air pollution affect heart-disease mortality<sup>9</sup> found no effects for ambient PM<sub>2.5</sub> from biomass burning in the United States. Nevertheless, as the authors point out, even if it is assumed that biomass burning and commercial and residential energy use do not contribute to mortality associated with heart disease, such energy use remains the largest factor for global mortality associated with air pollution overall, even though the total number of deaths declines.

Lelieveld and colleagues' next major finding is that agricultural sources are the second-largest contributor to global mortality from PM<sub>2.5</sub> — releases of ammonia from livestock and fertilizers lead to atmospheric formation of ammonium nitrate and sulfate particles. Agricultural sources are the leading source of mortality in the eastern United States, Russia, Turkey, Korea, Japan and Europe, contributing to more than 40% of the deaths in many European countries.

## ATMOSPHERIC SCIENCE

# The death toll from air-pollution sources

**Estimates of worldwide deaths associated with exposure to fine particles in atmospheric pollution provide some surprising results. The findings will guide future research and act as a wake-up call for policymakers. [SEE LETTER P.367](#)**

**MICHAEL JERRETT**

In this issue, Lelieveld *et al.*<sup>1</sup> (page 367) estimate the number of worldwide deaths each year caused by seven sources of air pollution. To do this, they used advanced global atmospheric-chemistry models, detailed country-level population and health data, and integrated exposure–response (IER) functions — statistical models that describe how mortality varies with exposure to fine particulate air pollution. The atmospheric-chemistry model allowed the researchers to attribute air pollution and premature deaths in different regions to emissions associated with various sectors of the economy.

More than 3.2 million deaths per year have been attributed<sup>2</sup> to exposure to outdoor particulate matter known as PM<sub>2.5</sub> — particles less than 2.5 micrometres in diameter, which can penetrate deep into the lungs and cause a wide range of health problems. Many parts of the United States and Europe have seen substantial improvements in air quality over recent decades as a result of regulatory interventions, and growing evidence<sup>3,4</sup> suggests that these improvements benefit public health. But

other regions, particularly countries in Asia with vast populations, continue to have poor air quality<sup>5</sup> (Fig. 1), with the emissions of several key pollutants expected to increase in the future<sup>6</sup>. The overlap of high pollution and large populations takes a huge toll on public health, but little is known about the pollution sources that are responsible for premature deaths.

Enter Lelieveld and colleagues. The authors' results are surprising and potentially important for protecting public health globally. First, they estimate that ambient PM<sub>2.5</sub> from commercial and residential energy sources contributes the most to premature deaths worldwide. These sources include solid fuel such as coal and biomass used for heating and cooking, local waste disposal and diesel generators. Such sources account for 32% of the premature deaths in China and 50–70% of those in India and other Asian nations.

The IER functions<sup>7</sup> that the authors used pool epidemiological exposure–response information for mortality associated with exposure to outdoor particles, emissions from biomass burning, and tobacco smoke (both from active smoking and second-hand exposure). For deaths attributable to stroke and





**Figure 1 | Burning waste in India.** Lelieveld *et al.*<sup>1</sup> estimate that fine particles generated from commercial and residential energy use, including waste burning, contribute the most to pollution-associated premature deaths globally, especially in India and other Asian countries.

This finding assumes that ammonium nitrate and sulfate have the same toxicity as other constituents of the atmospheric particle mixture. Some epidemiological studies<sup>10,11</sup> do indeed report adverse effects from these particles, but many toxicological data indicate that they have little biological potency at ambient levels<sup>10</sup>. The contradictory evidence for ammonium sulfate probably arises because these particles are often mixed with metals and other toxic components from coal or industrial sources<sup>11</sup>. It could therefore be that Lelieveld *et al.* overestimate the effects of particles from agricultural sources. The finding is highly valuable, however, because agriculture has generally not been seen as a major source of air pollution or premature death, and because it suggests that much more attention needs to be paid to agricultural sources, by both scientists and policymakers.

Third, the researchers find that traffic-related pollution accounts for about 20% of the deaths from PM<sub>2.5</sub> in the United States, the United Kingdom and Germany, but only 5% globally. The spatial resolution of their global assessment (which considers sub-areas of approximately 110 × 110 km) cannot capture the effects of finer-scale variation of traffic pollution. Other studies<sup>10,12</sup> have found that variation in pollution about 50–500 metres from the roadside correlates with mortality. Mounting evidence<sup>10</sup> also points to heightened effects on health and mortality from the components and reaction products of traffic emissions compared with other emission sources. Thus, the effects from traffic might be underestimated

by Lelieveld and colleagues. But the findings send out two crucial messages: traffic emissions remain a major source of premature death in Western countries even after extensive regulatory action, and the rapid rate of growth in traffic in many regions may well lead to increased pollution and more premature deaths in the near future.

Finally, the authors project a doubling of mortality from air pollution by 2050 on the basis of projected rates of increase in pollution and population levels. This projection should sound alarm bells for public-health agencies around the world. It also raises the question of which sources should be reduced in different regions. The answer depends on how much trust we put in the IER curve. Because the steep part of the curve is at lower levels of ambient PM<sub>2.5</sub>, large benefits can accrue from relatively small reductions in air pollution in cleaner regions, whereas the flatness of the curve at high levels necessitates large reductions in the polluted areas of Asia to achieve major health benefits<sup>13</sup>.

Lelieveld and colleagues' findings suggest that about 1 million lives could be saved every year by reducing ambient exposure to pollution. A further 3.54 million lives per year could be saved by lowering indoor exposures from similar sources<sup>2</sup>, mainly through changes in commercial and residential energy use. Incentivizing the use of cleaner fuels or of electricity for local energy needs would reduce mortality from both indoor and ambient PM<sub>2.5</sub> exposure and should be a priority in Asia and other regions that rely on solid fuels. For many parts

of the world, more research is needed if we are to understand the impacts of agricultural practices on air pollution and mortality, and especially to determine the toxicity of ammonium nitrate and sulfate emanating from this source. And in countries that already have low ambient levels of pollution, sizeable benefits can still be achieved by reducing emissions from fossil-fuel power plants and traffic. ■

**Michael Jerrett** is in the Department of Environmental Health Sciences, and at the Center for Occupational and Environmental Health, Fielding School of Public Health, University of California, Los Angeles, Los Angeles, California 90095.  
e-mail: mjerrett@ucla.edu

1. Lelieveld, J., Evans, J. S., Fnais, M., Giannadaki, D. & Pozzer, A. *Nature* **525**, 367–371 (2015).
2. Lim, S. S. *et al. Lancet* **380**, 2224–2260 (2012).
3. Pope, C. V. III, Ezzati, M. & Dockery, D. W. *N. Engl. J. Med.* **360**, 376–386 (2009).
4. Gauderman, W. J. *et al. N. Engl. J. Med.* **372**, 905–913 (2015).
5. Baumgartner, J. *et al. Proc. Natl Acad. Sci. USA* **111**, 13229–13233 (2014).
6. Wang, S. X. *et al. Atmos. Chem. Phys.* **14**, 6571–6603 (2014).
7. Burnett, R. T. *et al. Environ. Health Perspect.* **122**, 397–403 (2014).
8. Smith, K. R. *et al. Annu. Rev. Public Health* **35**, 185–206 (2014).
9. Thurston, G. D. *et al. Environ. Health Perspect.* (in the press).
10. Kelly, F. J. & Fussell, J. C. *Atmos. Environ.* **60**, 504–526 (2012).
11. Smith, K. R. *et al. Lancet* **374**, 2091–2103 (2009).
12. Hoek, G. *et al. Environ. Health* **28**, 12(1):43 (2013).
13. Apte, J. S., Marshall, J. D., Cohen, A. J. & Brauer, M. *Environ. Sci. Technol.* **49**, 8057–8066 (2015).

# Labelling and optical erasure of synaptic memory traces in the motor cortex

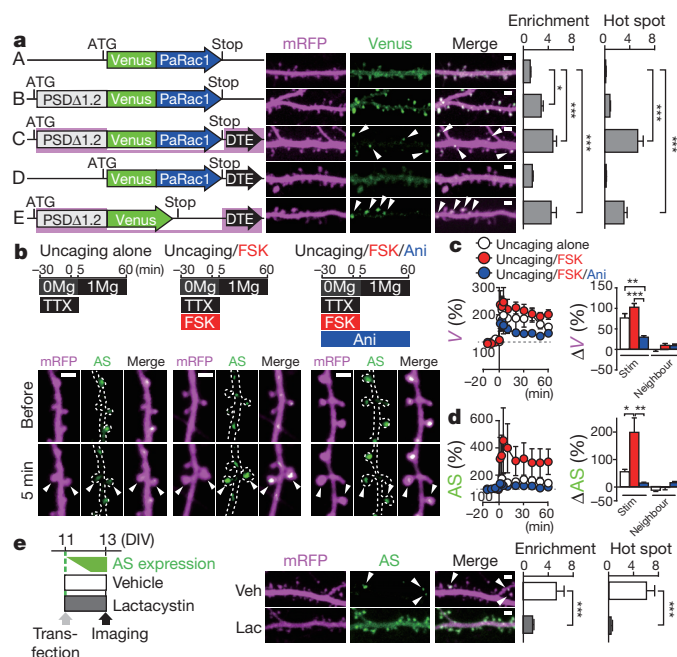
Akiko Hayashi-Takagi<sup>1,2</sup>, Sho Yagishita<sup>1,3</sup>, Mayumi Nakamura<sup>1</sup>, Fukutoshi Shirai<sup>1</sup>, Yi I. Wu<sup>4</sup>, Amanda L. Loshbaugh<sup>5,6</sup>, Brian Kuhlman<sup>5,6</sup>, Klaus M. Hahn<sup>5,7</sup> & Haruo Kasai<sup>1,3</sup>

**Dendritic spines are the major loci of synaptic plasticity and are considered as possible structural correlates of memory. Nonetheless, systematic manipulation of specific subsets of spines in the cortex has been unattainable, and thus, the link between spines and memory has been correlational. We developed a novel synaptic optoprobe, AS-PaRac1 (activated synapse targeting photoactivatable Rac1), that can label recently potentiated spines specifically, and induce the selective shrinkage of AS-PaRac1-containing spines. *In vivo* imaging of AS-PaRac1 revealed that a motor learning task induced substantial synaptic remodelling in a small subset of neurons. The acquired motor learning was disrupted by the optical shrinkage of the potentiated spines, whereas it was not affected by the identical manipulation of spines evoked by a distinct motor task in the same cortical region. Taken together, our results demonstrate that a newly acquired motor skill depends on the formation of a task-specific dense synaptic ensemble.**

Optogenetics is a powerful tool for controlling neuronal action potentials<sup>1,2</sup>, and has been used to demonstrate the crucial role of cell assemblies in representing memory traces<sup>3</sup>. However, owing to the limitations of spatial resolution of probes currently available, manipulation of individual dendritic spines, the major sites of excitatory synapses<sup>4–6</sup>, has been unfeasible, hindering the comprehensive understanding of synaptic reorganization during learning. Thus, for spine-specific light control, we took advantage of the structural properties of spines: the tight correlation between spine volume and function<sup>4–7</sup>. Because the prolonged activation of the small GTPase Rac1 induces spine shrinkage<sup>8–11</sup>, we used a photoactivatable form of Rac1 (PaRac1)<sup>12</sup> to induce spine shrinkage, which allowed us to control synaptic transmission with light. Moreover, since it has been suggested for a long time that the memory trace is allocated to specific neurons and spines of neurocircuits<sup>13,14</sup>, here we targeted PaRac1 to the activated synapses (activated synapse targeting PaRac1, AS-PaRac1) to establish a novel method, termed ‘synaptic optogenetics’, to visualize and manipulate the memory trace.

## AS-PaRac1 labels the potentiated spines

We first re-engineered the original PaRac1 construct<sup>12</sup> to optimize its properties for synaptic manipulation. Introduction of L514K and L531E mutations into the original construct markedly reduced the undesirable Rac1 background activity in the dark, as shown by isothermal titration calorimetry (ITC), the neuronal morphology, and co-immunoprecipitation (Extended Data Fig. 1a–c). Next, PaRac1 was fused with a deletion mutant of PSD-95 (PSDΔ1.2)<sup>15</sup>, which is known to concentrate at the postsynaptic site, but cannot bind with the major PDZ binding proteins, thus minimizing the undesirable effects of PSD-95 overexpression. An enrichment index, quantitative ratio of synaptic localization compared to that of the dendritic shaft (see Methods), supported the effective accumulation of PSD-PaRac1 to the synapse, especially at the tip of the spine (Fig. 1a, construct B), where it was highly co-localized with the endogenous PSD-95, but not with an axonal marker (Extended Data Fig.



**Figure 1 | Potentiation-dependent accumulation of AS-PaRac1 to the dendritic spines in hippocampal slice cultures.** **a**, Mapping for essential domains for the discrete distribution of the probe (arrowheads). Enrichment and hot spot index are plotted as arbitrary units. **b**, Representative images of single spine potentiations by glutamate uncaging (arrowheads) in the presence or absence of forskolin (FSK) and anisomycin (Ani). 0Mg, no Mg<sup>2+</sup>; 1Mg, 1 mM MgCl<sub>2</sub>. **c**, **d**, Time courses of spine head volume (*V*, **c**) and AS-PaRac1 accumulation (AS, **d**), both measured by fluorescence intensity. The mean change 60 min after uncaging in the stimulated or neighbouring spines. **e**, The effect of lactacystin on the discrete accumulation of AS-PaRac1 (arrowheads). DIV, days *in vitro*. Scale bars, 2 μm. Error bars represent s.e.m. Detailed information on statistical methods/results are described in Extended Data Table 1.

<sup>1</sup>Laboratory of Structural Physiology, Center for Disease Biology and Integrative Medicine, Faculty of Medicine, University of Tokyo, Bunkyo-ku, Tokyo 113-0033. <sup>2</sup>PRESTO, Japan Science and Technology Agency, 4-1-8 Honcho, Kawaguchi, Saitama 332-0012, Japan. <sup>3</sup>CREST, Japan Science and Technology Agency, 4-1-8 Honcho, Kawaguchi, Saitama 332-0012, Japan. <sup>4</sup>Center for Cell Analysis and Modeling, University of Connecticut Health Center, Farmington, Connecticut 06032, USA. <sup>5</sup>Lineberger Comprehensive Cancer Center, University of North Carolina, Chapel Hill, North Carolina 27599, USA. <sup>6</sup>Department of Biochemistry and Biophysics, University of North Carolina, Chapel Hill, North Carolina 27599, USA. <sup>7</sup>Department of Pharmacology, University of North Carolina, Chapel Hill, North Carolina 27599, USA.



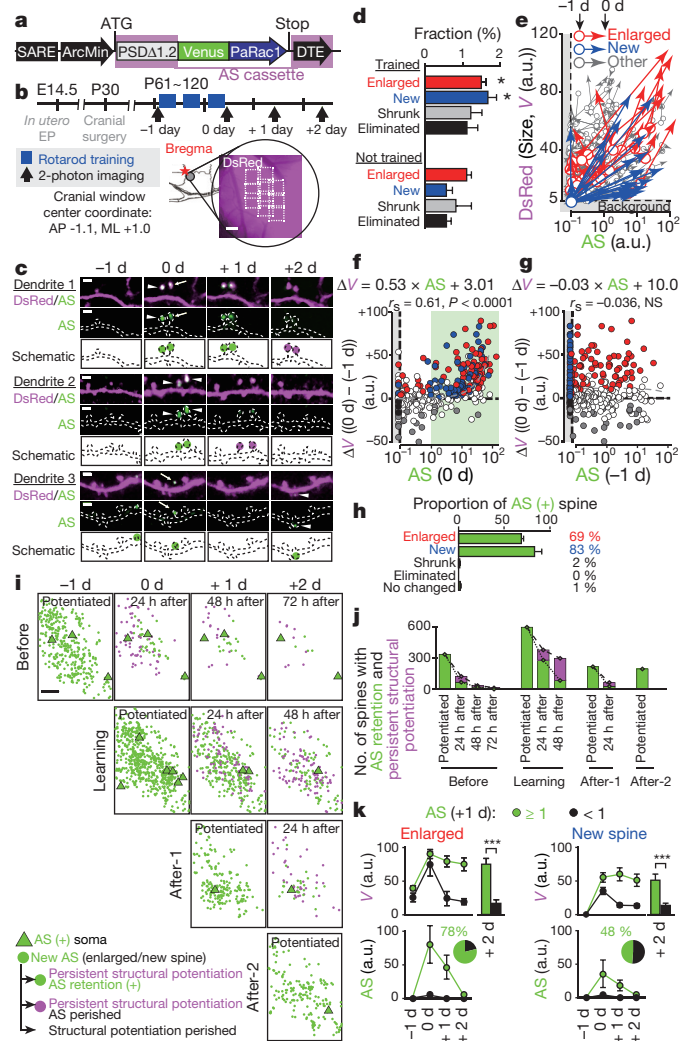
1d). Finally, for neuronal input specificity, we exploited the dendritic targeting element (DTE) of *Arc* mRNA<sup>16</sup>, which is selectively targeted and translated in activated dendritic segments in response to synaptic activation in an NMDA (N-methyl-D-aspartate) receptor-dependent manner<sup>17–19</sup>. Interestingly, PSD–PaRac1–DTE sparsely labelled spines (Fig. 1a, construct C, arrowheads). Quantification using a hot spot index (see Methods), which indicates how unevenly PaRac1 variants were distributed, suggested that both PSDΔ1.2 and DTE was necessary for this characteristic distribution (Fig. 1a, constructs C and E). Therefore, the combination of PSDΔ1.2 and DTE was termed as ‘AS (activated synapse targeting) cassette’, and the PaRac1 sequence flanked with the AS cassette was named AS–PaRac1 (Fig. 1a, construct C).

Next, we tried to unravel what this new synaptic probe labelled. Bicuculline, which increases neuronal excitation, robustly enhanced the number of AS–PaRac1-containing spines, and reduction of the hot spot index revealed that the distribution of AS–PaRac1 became relatively uniform upon bicuculline treatment. In contrast, the blockage of action potential by tetrodotoxin (TTX) decreased the accumulation of the probe, resulting in a reduction in the spine enrichment index of the probe (Extended Data Fig. 2a–d). Because these findings suggested that synaptic activation regulates the localization of AS–PaRac1, we hypothesized that AS–Rac1 accumulates in recently potentiated spines. Indeed, when AS–PaRac1 was co-transfected with SEP–GluA1, the synaptic incorporation marker for AMPA (α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid) receptor subunits GluR1 (refs 20, 21), the fluorescence signals of these two probes inside each spine were significantly correlated (Extended Data Fig. 2e, arrowheads). Furthermore, the protein synthesis-dependent potentiation during the single spine LTP protocol, which was elicited by glutamate uncaging and the adenylyl cyclase activator forskolin (FSK)<sup>22–24</sup>, induced the accumulation of AS–PaRac1 in the stimulated spines, while the protein synthesis-independent plasticity (glutamate uncaging alone) did not. Consistently, protein synthesis inhibitor anisomycin abolished the FSK-induced AS–PaRac1 accumulation (Fig. 1b–d). No increase was observed in AS–PaRac1 fluorescence in the neighbouring spines, indicating that AS–PaRac1 accumulation was restricted to the stimulated spine (Fig. 1d). The DTE sequence was necessary for activity-dependent AS–PaRac1 accumulation (Extended Data Fig. 2f, g), supporting that locally translated AS–PaRac1, unlike somatically translated AS–PaRac1, was preferentially recruited to enlarged spines. PaRac1 did not exhibit uneven distribution unless the construct contained the PSD-95 domain (Fig. 1a, construct D). Because PSD-95 is rapidly degraded by proteasomes<sup>25</sup>, we examined the effect of the proteasome inhibitor lactacystin and found that it completely disrupted the unique distribution of the probe (Fig. 1e). Taken together, we concluded that AS–PaRac1 is a probe that specifically labels the enlarged and newly generated spines (see Extended Data Fig. 3 for detailed cellular mechanisms), which are referred to as the ‘structurally potentiated spine’, and the potentiation labelled by AS–PaRac1 is described as ‘potentiated spine’ hereafter.

### Spine labelling by AS–PaRac1 *in vivo*

To characterize this probe *in vivo*, we used the rotarod training as the model of motor learning. Because motor learning is impaired in *Arc* knockout mice<sup>26</sup>, we assumed that the induction of AS–PaRac1 by the *Arc* promoter<sup>27</sup> would enhance specific labelling during learning-induced potentiation. Arc::AS–PaRac1 was delivered to the cortical layer II/III of the primary motor cortex (M1), where a robust reorganization of neuronal circuits is induced upon motor learning<sup>28–31</sup>. Cranial window surgery for two-photon imaging was performed based on the stereotaxic coordinates of the previous functional mapping for the hind limb area<sup>32</sup>. Spine volume and AS–PaRac1 fluorescence was compared quantitatively before and after training (Fig. 2a–e). Consistent with previous findings<sup>33,34</sup>, even in the training-free period, a substantial number of spines ‘spontaneously’ under-

went structural potentiation (formation or enlargement of spines; see the definition in Extended Data Fig. 4a), but the trained mice exhibited significantly more structural potentiation compared with the non-trained mice (Fig. 2d). Notably, synaptic fluorescence of AS–PaRac1 just after training (0 day) strongly correlated with the change in spine size upon training (Fig. 2f). It is unlikely that the accumulation of AS–PaRac1 caused the potentiation or labelled the spines primed for potentiation such as for the ‘tagged synapse’<sup>35,36</sup>, because the initial quantity of AS–PaRac1 before learning (–1 day) did not correlate with the change in spine size after learning (Fig. 2g). Analysis of AS–PaRac1 puncta in the dendritic shaft suggested that the majority



**Figure 2 | Spatiotemporal dynamics of AS–PaRac1 labelling *in vivo* during the rotarod task.** **a**, Schematic of the Arc promoter<sup>27</sup>-driven AS–PaRac1.

**b**, Experimental design. EP, electroporation; DsRed, *Drosophila* sp. red fluorescent protein. **c**, Images of spine formation (arrows) and spine enlargement (arrowheads). Green circles, AS–PaRac1; magenta circles, spines that initially acquired AP–PaRac1, but lost it afterward, but the structural change was persistent. **d**, Fraction of structural change of spines. **e–g**, Quantification of spine size and AS–PaRac1 (**e**). Size measured by fluorescence intensity, arbitrary units (a.u.). Relationship between AS–PaRac1 and volume change ( $\Delta V$ ) after (0 day, **f**) and before (–1 day, **g**) learning. **h**, Percentage of AS–PaRac1-containing spines (AS–PaRac1  $\geq 1$  a.u., area shaded green in **f**). **i**, Mapping of AS–PaRac1. Potentiations before, just after, 1 day after, and 2 days after learning are separately depicted as ‘Before’, ‘Learning’, ‘After-1’, and ‘After-2’, respectively. **j**, Retention of AS–PaRac1 (green) or structural potentiation (magenta). **k**, Trajectory of spine size and AS–PaRac1 intensities of the structurally potentiated spines. Scale bars, 2  $\mu\text{m}$  for **c**; 200  $\mu\text{m}$  for **b** and **i**. Error bars represent s.e.m. NS, not significant.

of AS-PaRac1 signal was located in the dendritic spines, and the labelling of shaft synapses was negligible (Extended Data Fig. 5). When we set the threshold of AS-PaRac1 at 1 a.u. (Fig. 2f, green shaded area, 0 day), AS-PaRac1 detected spine formation and enlargement with sensitivities of  $83 \pm 7.9\%$  and  $69 \pm 3.0\%$  (mean  $\pm$  standard error of the mean), respectively (Fig. 2h), whereas labelling in other spine types was  $2.3 \pm 0.12\%$ . Since Arc::AS-PaRac1 was induced only in the AS-PaRac1-positive neuron, the labelling properties in the AS-PaRac1-positive neuron were also calculated. The sensitivities for formation and enlargement were  $94 \pm 2.7\%$  and  $95 \pm 4.8\%$ , respectively, while false labelling in other spine types was  $12.9 \pm 4.2\%$  (306 spines, 6 AS-PaRac1-positive neurons in 3 mice). Therefore, AS-PaRac1 is a reliable marker of the potentiated spines *in vivo*.

Next, we performed wide-view mapping of task-evoked potentiation using this probe (Fig. 2i, learning period), and we found that the task-evoked potentiation was elicited in  $2.3 \pm 0.13\%$  of spines and  $16.4 \pm 2.8\%$  of neurons in the imaged area. We tracked an almost whole image of neurons (Extended Data Fig. 4b) and confirmed that when a spine was labelled by AS-PaRac1, its parental soma also expressed AS-PaRac1 (6 AS-PaRac1-positive somata). Consistently, we could not find AS-PaRac1-positive spines in AS-PaRac1-negative soma (46 negative somata). Thus, the counting of AS-PaRac1 puncta per AS-PaRac1-positive neurons could be approximated, which suggested that  $14.7 \pm 0.01\%$  of spines contained AS-PaRac1 in the AS-PaRac1-positive neurons, implying that upon motor learning, a substantial remodelling of spines (14.7%) was evoked in a small neuronal population (16.4%) in layer II/III (Extended Data Fig. 4d for detailed calculation). Similarly, a substantial remodelling was also observed in a small population of layer V neurons (Extended Data Fig. 4d).

To characterize the synaptic retention of AS-PaRac1 for photoactivation experiments *in vivo*, the individual spines that acquired AS-PaRac1 were tracked, and were separately schematized from the day of AS-PaRac1 appearance (Fig. 2c, i). We noticed that persistence of synaptic AS-PaRac1 and the structural potentiation markedly varied among spines: some were preserved beyond 1 day after training (Fig. 2c, dendrite no. 1), while others disappeared (Fig. 2c, dendrites no. 2 and 3). Importantly, the structural potentiation and AS-PaRac1 labelling triggered during the 'learning' period were more likely to be preserved than those triggered during the training-free period (Fig. 2j, 'Before' and 'After-1 (potentiation 1 day after learning)'). Consistently, longitudinal imaging of the structurally potentiated spines revealed that the majority of those retaining AS-PaRac1 for 24 h maintained structural potentiation for at least 48 h (Fig. 2k, green trace), whereas the structurally potentiated spines lacking AS-PaRac1 retention returned to the pre-potentiated state (Fig. 2k, black trace). Such AS-PaRac1 retention might be maintained by reverberation of learning-activated neuronal circuits, because AS-PaRac1 was only expressed in Arc-expressing neurons, in which the persistent activation helps to maintain plastic changes in the neocortex<sup>26,37</sup>.

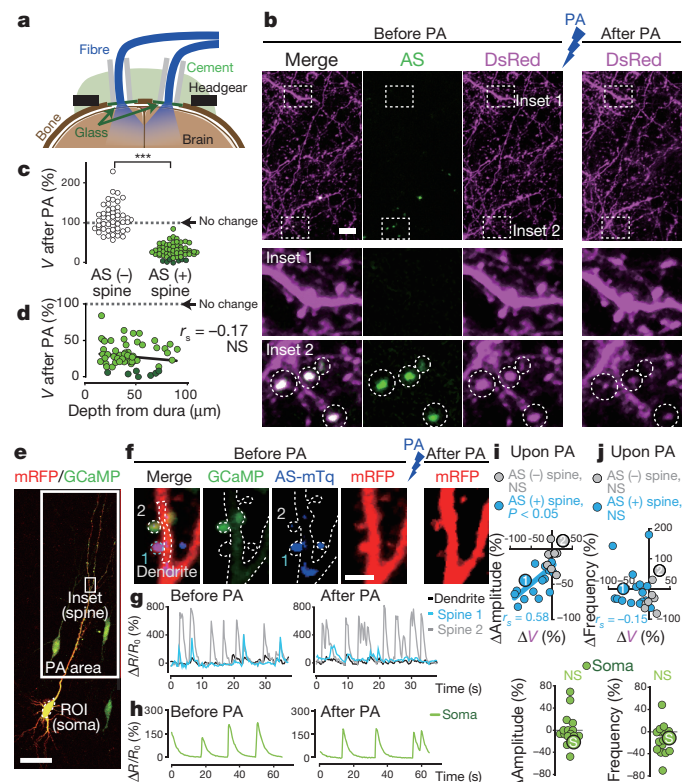
### Selective spine shrinkage by AS-PaRac1

Consistent with the previous findings that prolonged Rac1 activation induces spine shrinkage<sup>8–11</sup>, we found that low-frequency photoactivation elicited spine shrinkage (Extended Data Fig. 6). Intriguingly, the spine shrinkage was significantly more robust when the AS-PaRac1 construct was driven by the Arc promoter compared with the constitutive promoter CAG. Arc expression is increased by persistent neuronal activity<sup>26</sup>, which induces the chronic activation of endogenous Rac1, possibly contributing to the robust spine shrinkage by Arc::AS-PaRac1. Photoactivation-induced spine shrinkage was Rac1-dependent, because deletion of Rac1 from AS-PaRac1 while keeping other domains intact within AS-PaRac1 (Arc::PSDΔ1.2-LOV-DTE) completely disrupted the shrinkage effect (Extended Data Fig. 6). To achieve spine shrinkage in a large cortical area *in vivo*, bilateral optical fibres were placed onto the cranial window (Fig. 3a and Extended Data Fig. 7). Low-frequency pulsed photoacti-

vation triggered shrinkage specifically in the AS-PaRac1-containing spines (Fig. 3b, c). The effect of photoactivation was comparable at least within 100  $\mu\text{m}$  from the dura, suggesting that spines in layer I, at least, were affected by photoactivation (Fig. 3d). Photoactivation-induced spine shrinkage was accompanied by functional depotentiation, which was demonstrated by the excitatory postsynaptic calcium transients: the extent of spine shrinkage correlated with the decrease in amplitude, but not with the decrease in frequency (Fig. 3f–j). Spine shrinkage and the subsequent functional changes were spine-specific, but not branch- or cell-wide, because spine shrinkage was not triggered in neighbouring AS-PaRac1-negative spines, and the calcium transient was not affected either in the neighbouring spines or in the soma (Fig. 3f–j; Extended Data Figs 6 and 7b).

### Optical erasure of acquired skills

To demonstrate the effect of spine shrinkage for learning, mice were bilaterally injected with the adeno-associated virus (AAV) 5 that encompassed layers I to V (Extended Data Fig. 7f). Mice were divided into two groups: animals in the first group were transfected with monomeric red fluorescent protein (mRFP) alone as a control, and the second group was transfected with AS-PaRac1 and mRFP. Both groups exhibited significantly better motor performance after training, but only the performance of the AS-PaRac1 group was disturbed by photoactivation (protocol 1, Fig. 4a, b), and the extent of learning disruption induced by photoactivation (photoactivation



**Figure 3 | Selective shrinkage of AS-PaRac1-containing spines upon photoactivation (PA).** **a**, Illustration of photoactivation. **b**, Images of the hind limb regions of cortices. SARE::AS-PaRac1 and CAG::mRFP were transduced by *in utero* EP (**a–d**). **c**, Spine size following photoactivation. Dark green circles are eliminated spines. **d**, The effect of cortical depth on photoactivation-induced spine shrinkage. **e–j**, Hippocampal neuron cultures were co-transfected with GCaMP6s, AS-PaRac1-mTurquoise, and mRFP. Changes in the GCaMP/mRFP ratio ( $\Delta R$ ) in synapse (**g**) and soma (**h**) were traced. **i**, **j**, Relationships between  $\Delta V$  and  $\Delta$ Amplitude (**i**), or  $\Delta$ Frequency (**j**) upon photoactivation. Circles 1, 2, and S correspond to spine no. 1, no. 2, and the soma in **g** and **h**. Scale bars, 5  $\mu\text{m}$  for **b**; 50  $\mu\text{m}$  for **e**; 2  $\mu\text{m}$  for **f**.



effect) negatively correlated with the extent of training-evoked improvement (learning attainment) (Fig. 4f). In contrast, there was neither disruption of acquired learning nor a correlation between the effects of the training and photoactivation in the control group. Since photoactivation did not affect the running speed of the identical cohort used in Fig. 4b, it is unlikely that photoactivation disturbed the general motor performance (Extended Data Fig. 8).

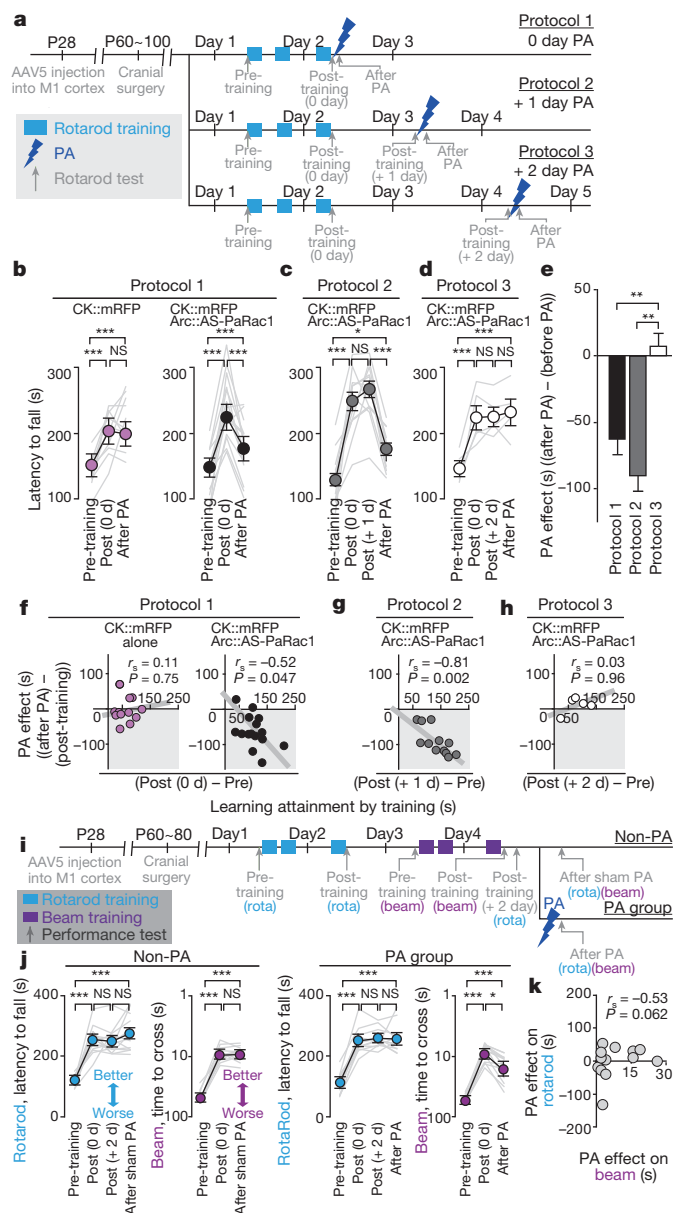
Photoactivation disrupted the acquired learning even 1 day after learning (protocol 2; Fig. 4c, g), when the majority of learning-evoked spines contained AS-PaRac1 (Fig. 2k). In contrast, photoactivation treatment 2 days after learning (protocol 3), when both the number of AS-PaRac1-containing spines and the intensity of AS-PaRac1 labelling were markedly decreased (Fig. 2k; Extended Data Fig. 4), failed to disrupt acquired learning (Fig. 4d, h). Owing to daily spontaneous potentiation, a comparable number of spines contained AS-PaRac1 in both protocol 2 and protocol 3 (Extended Data Fig. 4c). Nonetheless, only protocol 2 disrupted the acquired skill, suggesting that the learning-evoked spine potentiation visualized by AS-PaRac1 (at +1 day), but not spontaneous potentiation (at +2 day), accounted for the cortical memory traces.

To demonstrate the task-specific role of synaptic ensembles, mice injected with AS-PaRac1-expressing AAV into the bilateral M1 were subjected to a dual task protocol. Mice sequentially learned two distinct hind limb tasks: the rotarod and the beam tasks in the first and second sets of 2 days, respectively (Fig. 4i). We performed the photoactivation on day 4, because the majority of the rotarod-evoked AS-PaRac1 puncta diminished by this time point (Fig. 2k). We confirmed that these two tasks evoked a comparable number of spine potentiation (Extended Data Fig. 7c). While learning performance in the beam task was not disrupted by the sham photoactivation treatment (fibre was inserted, but no illumination was performed), photoactivation disrupted the acquired performance in the beam task, without affecting the rotarod performance (Fig. 4j). We found no correlation between the effect of photoactivation in the rotarod and the beam task, which implies that synaptic ensembles recruited by each task did not overlap (Fig. 4k).

### Task-specific synaptic ensemble

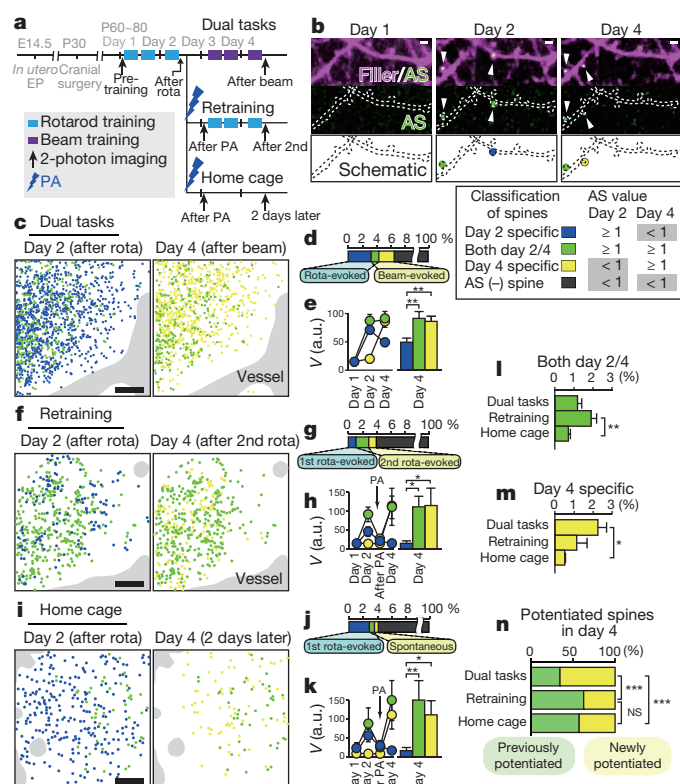
To visualize the synaptic ensembles formed during dual task learning, mice were sparsely labelled with AS-PaRac1, and were also subjected to the dual task protocol described before (Fig. 5a, dual task). AS-PaRac1 puncta were classified on the basis of time of emergence (Fig. 5b), schematized for the rotarod task potentiation (day 2 specific) as blue dots, for the beam task potentiation (day 4 specific) as yellow dots, and for the continuous potentiation for both periods (both day 2 and 4) as green dots. Interestingly, more than half of the beam-evoked potentiation were new ones (Fig. 5n), which were not potentiated previously in the rotarod task (yellow, Fig. 5c–e). Taken together with the behavioural data (Fig. 4i–k), we have demonstrated that the two learning tasks induced the potentiation of distinct synaptic ensembles.

Finally, we examined whether the same spines are potentiated by the same task. Mice were divided into 2 groups (Fig. 5a). The first group was subjected to the rotarod task in the first 2 days, which was followed by the shrinkage of the learning-evoked potentiation by photoactivation, and then the identical rotarod task was re-trained (re-training condition). The second group was subjected to the rotarod task and subsequent photoactivation, and mice were not trained for another 2 days (home cage condition). We found that the majority of the optically shrunk spines returned to their previously potentiated size after re-training, while the degree of re-potentiation was significantly lower in the home cage group, suggesting that re-training induced the re-potentiation of the same subset of spines (Extended Data Fig. 7d, e). Mice assigned to the dual task protocol were also compared, highlighting the difference in the potentiation patterns among the groups during the last 2 days



**Figure 4 | Erasure of acquired learning by the photoactivation of spines labelled with AS-PaRac1.** **a**, Experimental design (see Extended Data Fig. 9). **b–d**, Mice, which received AAV infection of SARE::AS-PaRac1, were allocated to protocols 1 (**b**), 2 (**c**), or 3 (**d**). An average of three trials of each mouse was used as the task performance (grey line). **e**, The critical period of photoactivation to erase acquired skills. **f–h**, Relationship between the effect of photoactivation and learning attainments. **i**, Experimental design. **j**, Performance trajectory of each skill. **k**, No correlation between photoactivation effect on acquired rotarod performance and that of beam task. Error bars represent s.e.m.

(Fig. 5c–n). Contrary to the dual task group, spines potentiated during the first rotarod training were more likely to be re-potentiated after the second rotarod training in the re-training group (green, Fig. 5f–h, l, n), while re-potentiation was significantly less prominent in mice that did not perform the re-training task (home cage group) (Fig. 5i–k, l, n; Extended Data Fig. 7d, e). Furthermore, newly potentiated spines, which were not potentiated in the first 2 days, were less abundant in the re-training and home cage groups compared with the dual task group (yellow, Fig. 5m, n). These findings suggest that reorganization of distinct synaptic ensembles is specific for each learning task.



**Figure 5 | Visualization of synaptic ensembles for distinct learning tasks.** **a**, Experimental design. Arc::AS-PaRac1 and CAG::mRFP (filler) was transduced by *in utero* electroporation. **b**, Images of dendrites upon learning. AS-PaRac1 puncta are colour-coded based on its appearance and duration. Identical colour codes are used in **c–n**. **c, f, i**, Wide view mapping of AS-PaRac1. **d, g, j**, The fraction of each spine type. **e, h, k**, The trajectory of spine size ( $V$ ). **l, m**, Differential spine potentiation in each condition. **n**, The proportions of newly potentiated spines. Scale bars, 2  $\mu$ m for **b**; 50  $\mu$ m for **c, f** and **i**. Error bars represent s.e.m.

## Discussion

Current models of learning and memory suggest that structural plasticity of spines is the underlying mechanism of information storage in the brain. Nonetheless, clear visualization of spine structure *in vivo* requires the sparse labelling of neurons, and analysis of structural changes in spines is very laborious. In contrast, the AS-PaRac1 signal appears as fluorescence puncta, which allows the detection of potentiated spines far more easily, even at high transfection condition. Moreover, the role of potentiated spines can be directly assessed with photoactivation during behavioural examinations. In this study, we showed that photoactivation of the bilateral M1 cortex disrupted the acquired motor skill. We estimated the number of learning-evoked neurons affected by photoactivation was approximately 4,700 neurons based on the following calculation: (a)  $\times$  (b)  $\times$  (c)  $\times$  (d)  $\times$  (e), in which (a) represents the density of neurons in the neocortex,  $9.2 \times 10^4/\text{mm}^3$  (ref 38); (b) the photoactivated area, fibre core diameter = 500  $\mu$ m,  $0.4 \text{ mm}^2/\text{bilateral}$ ; (c) the thickness of cortical layers (II–V) that were infected with AAV, 0.8 mm; (d) AAV infection efficiency, 80% (Extended Data Fig. 7f); (e) the percentage of AS-PaRac1-positive neurons upon learning, 20% (Extended Data Fig. 4d). On the other hand, due to the limitations of light transmission, the majority of the shrunk spines resided in layer I (up to 100  $\mu$ m from the dura). The minimal number of learning-evoked spines illuminated by the optical fibre was roughly 410,000 spines in the bilateral M1 cortex based on the following calculation: (d)  $\times$  (f)  $\times$  (g)  $\times$  (h), in which (f) represents the density of excitatory synapses in the mouse neocortex,  $6.4 \times 10^8/\text{mm}^3$  (ref 38); (g) learning-evoked potentiation,

approximately 2% of the spines in this area (Extended Data Fig. 4); (h) brain volume that received photoactivation:  $0.4 \text{ mm}^2$  of photoactivation area  $\times$   $0.1 \text{ mm}$  of depth =  $0.04 \text{ mm}^3$ ). In the layer I, corticocortical feedback projections mediating top-down influences are concentrated, which strongly excite a subpopulation of pyramidal neurons<sup>39</sup>. Learning-evoked changes in neuronal ensembles via the synaptic reorganization of the M1 cortex directly predict future task performance<sup>40</sup>. As nonlinear information integration primarily occurs in the tuft of dendrites in behaving animals<sup>41</sup>, and activation of several spines in the tuft is sufficient to initiate NMDA spikes for action potential generation<sup>42</sup>. Thus, the shrinkage of potentiated spines in our study (410,000 spines in the dendritic tufts of 4,700 neurons) would be reasonably expected to disrupt the learning-evoked substantial remodelling in a specific neuronal population. Formation of the dense connections in a small neuronal ensemble may be consistent with the formation of functional neuronal clusters in the motor cortex after learning<sup>43</sup>. Thus, synaptic optogenetics might be a powerful tool to uncover the mechanism of synaptic plasticity and its relationships with subsequent behavioural manifestations.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 6 May 2014; accepted 3 August 2015.

Published online 9 September 2015.

- Bernstein, J. G. & Boyden, E. S. Optogenetic tools for analyzing the neural circuits of behavior. *Trends Cogn. Sci.* **15**, 592–600 (2011).
- Tye, K. M. & Deisseroth, K. Optogenetic investigation of neural circuits underlying brain disease in animal models. *Nature Rev. Neurosci.* **13**, 251–266 (2012).
- Liu, X. *et al.* Optogenetic stimulation of a hippocampal engram activates fear memory recall. *Nature* **484**, 381–385 (2012).
- Holtmaat, A. & Svoboda, K. Experience-dependent structural synaptic plasticity in the mammalian brain. *Nature Rev. Neurosci.* **10**, 647–658 (2009).
- Kasai, H., Fukuda, M., Watanabe, S., Hayashi-Takagi, A. & Noguchi, J. Structural dynamics of dendritic spines in memory and cognition. *Trends Neurosci.* **33**, 121–129 (2010).
- Yuste, R. Dendritic spines and distributed circuits. *Neuron* **71**, 772–781 (2011).
- Murakoshi, H. & Yasuda, R. Postsynaptic signaling during plasticity of dendritic spines. *Trends Neurosci.* **35**, 135–143 (2012).
- Luo, L. *et al.* Differential effects of the Rac GTPase on Purkinje cell axons and dendritic trunks and spines. *Nature* **379**, 837–840 (1996).
- Tashiro, A., Minden, A. & Yuste, R. Regulation of dendritic spine morphology by the rho family of small GTPases: antagonistic roles of Rac and Rho. *Cereb. Cortex* **10**, 927–938 (2000).
- Hayashi-Takagi, A. *et al.* Disrupted-in-Schizophrenia 1 (DISC1) regulates spines of the glutamate synapse via Rac1. *Nature Neurosci.* **13**, 327–332 (2010).
- Hayashi-Takagi, A. *et al.* PAKs inhibitors ameliorate schizophrenia-associated dendritic spine deterioration *in vitro* and *in vivo* during late adolescence. *Proc. Natl Acad. Sci. USA* **111**, 6461–6466 (2014).
- Wu, Y. I. *et al.* A genetically encoded photoactivatable Rac controls the motility of living cells. *Nature* **461**, 104–108 (2009).
- Redondo, R. L. & Morris, R. G. Making memories last: the synaptic tagging and capture hypothesis. *Nature Rev. Neurosci.* **12**, 17–30 (2011).
- Rogerson, T. *et al.* Synaptic tagging during memory allocation. *Nature Rev. Neurosci.* **15**, 157–169 (2014).
- Arnold, D. B. & Clapham, D. E. Molecular determinants for subcellular localization of PSD-95 with an interacting K<sup>+</sup> channel. *Neuron* **23**, 149–157 (1999).
- Kobayashi, H., Yamamoto, S., Maruo, T. & Murakami, F. Identification of a *cis*-acting element required for dendritic targeting of activity-regulated cytoskeleton-associated protein mRNA. *Eur. J. Neurosci.* **22**, 2977–2984 (2005).
- Steward, O., Wallace, C. S., Lyford, G. L. & Worley, P. F. Synaptic activation causes the mRNA for the IEG Arc to localize selectively near activated postsynaptic sites on dendrites. *Neuron* **21**, 741–751 (1998).
- Steward, O. & Worley, P. F. Selective targeting of newly synthesized Arc mRNA to active synapses requires NMDA receptor activation. *Neuron* **30**, 227–240 (2001).
- Korb, E. & Finkbeiner, S. Arc in synaptic plasticity: from gene to behavior. *Trends Neurosci.* **34**, 591–598 (2011).
- Makino, H. & Malinow, R. Compartmentalized versus global synaptic plasticity on dendrites controlled by experience. *Neuron* **72**, 1001–1011 (2011).
- Zhang, Y., Cudmore, R. H., Lin, D. T., Linden, D. J. & Huganir, R. L. Visualization of NMDA receptor-dependent AMPA receptor synaptic plasticity *in vivo*. *Nature Neurosci.* **18**, 402–407 (2015).
- Govindarajan, A., Israely, I., Huang, S. Y. & Tonegawa, S. The dendritic branch is the preferred integrative unit for protein synthesis-dependent LTP. *Neuron* **69**, 132–146 (2011).
- Tanaka, J. *et al.* Protein synthesis and neurotrophin-dependent structural plasticity of single dendritic spines. *Science* **319**, 1683–1687 (2008).



24. Harvey, C. D. & Svoboda, K. Locally dynamic synaptic learning rules in pyramidal neuron dendrites. *Nature* **450**, 1195–1200 (2007).
25. Colledge, M. *et al.* Ubiquitination regulates PSD-95 degradation and AMPA receptor surface expression. *Neuron* **40**, 595–607 (2003).
26. Ren, M., Cao, V., Ye, Y., Manji, H. K. & Wang, K. H. Arc regulates experience-dependent persistent firing patterns in frontal cortex. *J. Neurosci.* **34**, 6583–6595 (2014).
27. Kawashima, T. *et al.* Synaptic activity-responsive element in the Arc/Arg3.1 promoter essential for synapse-to-nucleus signaling in activated neurons. *Proc. Natl Acad. Sci. USA* **106**, 316–321 (2009).
28. Yu, X. & Zuo, Y. Spine plasticity in the motor cortex. *Curr. Opin. Neurobiol.* **21**, 169–174 (2011).
29. Masamizu, Y. *et al.* Two distinct layer-specific dynamics of cortical ensembles during learning of a motor task. *Nature Neurosci.* **17**, 987–994 (2014).
30. Peters, A. J., Chen, S. X. & Komiyama, T. Emergence of reproducible spatiotemporal activity during motor learning. *Nature* **510**, 263–267 (2014).
31. Cichon, J. & Gan, W. B. Branch-specific dendritic  $Ca^{2+}$  spikes cause persistent synaptic plasticity. *Nature* **520**, 180–185 (2015).
32. Hira, R. *et al.* Transcranial optogenetic stimulation for functional mapping of the motor cortex. *J. Neurosci. Methods* **179**, 258–263 (2009).
33. Yang, G., Pan, F. & Gan, W. B. Stably maintained dendritic spines are associated with lifelong memories. *Nature* **462**, 920–924 (2009).
34. Fu, M., Yu, X., Lu, J. & Zuo, Y. Repetitive motor learning induces coordinated formation of clustered dendritic spines *in vivo*. *Nature* **483**, 92–95 (2012).
35. Frey, U. & Morris, R. G. Synaptic tagging and long-term potentiation. *Nature* **385**, 533–536 (1997).
36. Okada, D., Ozawa, F. & Inokuchi, K. Input-specific spine entry of soma-derived Vesl-1S protein conforms to synaptic tagging. *Science* **324**, 904–909 (2009).
37. Wang, K. H. *et al.* *In vivo* two-photon imaging reveals a role of Arc in enhancing orientation specificity in visual cortex. *Cell* **126**, 389–402 (2006).
38. Schüz, A. & Palm, G. Density of neurons and synapses in the cerebral cortex of the mouse. *J. Comp. Neurol.* **286**, 442–455 (1989).
39. Cauller, L. Layer I of primary sensory neocortex: where top-down converges upon bottom-up. *Behav. Brain Res.* **71**, 163–170 (1995).
40. Laubach, M., Wessberg, J. & Nicolelis, M. A. Cortical ensemble activity increasingly predicts behaviour outcomes during learning of a motor task. *Nature* **405**, 567–571 (2000).
41. Xu, N. L. *et al.* Nonlinear dendritic integration of sensory and motor input during an active sensing task. *Nature* **492**, 247–251 (2012).
42. Larkum, M. E., Nevian, T., Sandler, M., Polsky, A. & Schiller, J. Synaptic integration in tuft dendrites of layer 5 pyramidal neurons: a new unifying principle. *Science* **325**, 756–760 (2009).
43. Hira, R. *et al.* Spatiotemporal dynamics of functional clusters of neurons in the mouse motor cortex during a voluntary movement. *J. Neurosci.* **33**, 1377–1390 (2013).

**Acknowledgements** We thank H. Bito and H. Okuno for the generous gift of the Arc promoter; F. Murakami for the information about Arc 3' UTR; M. Yuzaki, K. Inokuchi, and K. Fox for discussions. This research was supported by Grants-in-Aid from the Ministry of Education, Culture, Sports, Science, and Technology (MEXT, Japan; No. 2000009 to H.K. and No. 26221011 to H.K. and A.H.-T., No. 23689055 and No. 24116003 to A.H.-T.), the PRESTO program (JST) to A.H.-T., the brain/MIND and SICP projects from Japan Agency for Medical Research and Development (AMED) to H.K., the National Institutes of Health grant GM102924 to K.M.H., NS071216 to Y.I.W. and the Research Grant from the Human Frontier Science Program to H.K., K.M.H. and B.K.

**Author Contributions** A.H.-T., S.Y., M.N., and F.S. conducted the experiments. Y.I.W., A.L.L., K.M.H., and B.K. provided technical support for the development of PaRac1. A.H.-T. and H.K. designed the study and wrote the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.H.-T. (hayashi888@m.u-tokyo.ac.jp) or H.K. (hkasai@m.u-tokyo.ac.jp).

## METHODS

**Ethical considerations.** The use and care of animals in this study followed the guidelines of the Animal Experimental Committee of the Faculty of Medicine at the University of Tokyo.

**Plasmid construction and transfection.** Mutagenesis and deletion of cDNA were conducted based on previously described methods<sup>10</sup>. Briefly, L514K and L531E mutations of the LOV2 domain were introduced with the following primers (mutations underlined): 5'-cttattgggttcagaaggtgaactgagcatg-3', 5'-gagaggggagtcagtgagattaagaaacagcag-3', and with their corresponding complementary primers. PSD-95( $\Delta$ PDZ1.2) was generated by deleting the nucleotides (nts) 250 to 993 based on the numbering of NM\_019621. The DTE sequence of Arc mRNA was cloned from the 1st strand cDNA generated from the frontal cortex of postnatal day 50 (P50) Sprague–Dawley rats with the following primers (HindIII underlined): 5'-atgataagcttcggctccatgactcagccatgcc-3' and 5'-atgaatagcttagacagcagcagttaccaacagc-3'. The generated amplicon, which corresponded to 2036–2699 nts based on the numbering of NM\_019361, was subcloned immediately downstream of the stop codon of PaRac1.

**Isothermal titration calorimetry (ITC).** ITC for examining the affinity of PaRac1 to the CRIB domain of PAK1 in the lit and dark states was carried out as described previously<sup>12</sup>.

**PaRac1 pull-down assay.** PaRac1 variants were transfected into HEK293 cells by lipofection (Lipofectamine 2000; Invitrogen, Carlsbad, CA), and the cells were divided into lit and dark groups. The cells in the lit group were illuminated with a white fluorescent lamp (1.5 W for a 10-cm dish,  $19 \pm 1.0$  mW cm<sup>-2</sup>) for 10 min before cell lysis, and the subsequent immunoprecipitation was performed in continuous light illumination until the final wash step of protein precipitants. Cells in the dark group were manipulated under a yellow fluorescence lamp, which excluded light at the wavelengths below 500 nm to avoid photoactivation. Cells were lysed in a lysis buffer (150 mM NaCl, 50 mM Tris-HCl pH 7.5, 1% Triton-X (v/v), 10 mM NaF, 10% glycerol (v/v), 1 mM EDTA, and protein inhibitor cocktail (Complete; Roche Diagnostics)). Lysates were sonicated intermittently on the mixture of ice and water, and cell debris was cleared by centrifugation. The soluble fraction was incubated with an anti-GFP antibody (D253-3; MBL, Nagoya, Japan), followed by co-precipitation with Protein G Sepharose (GE Healthcare, Little Chalfont, UK). The precipitate was immunoblotted with an anti-PAK1 antibody (no. 2602; Cell Signaling, Beverly, MA). Signal intensity of each band (net signal after subtracting the background signal, which was obtained from the region adjacent to the band) was measured using the ImageJ software (National Institutes of Health, Bethesda, MD).

**Immunofluorescence.** Cell staining was performed as described previously<sup>10</sup>. Briefly, dissociated rat cortical neurons at 21 days *in vitro* (DIV) were fixed with 4% paraformaldehyde (PFA) for 30 min at room temperature. Mice were euthanized after the behavioural analyses, and their brains were perfusion-fixed with 4% PFA and sectioned coronally to obtain 150- $\mu$ m thick sections. Fixed samples were then permeabilized with Perm/Blocking buffer (2.5% normal goat serum (v/v) in phosphate-buffered saline [PBS] with 0.3% Triton X-100 (v/v)) for 1 h at room temperature. Samples were incubated for 24 h at 4 °C with the following primary antibodies: anti-phospho-neurofilament (SMI-31; Merck KGaA, Darmstadt, Germany), axonal marker; anti-PSD-95 (6G6; Abcam, Cambridge, UK); anti-Emx1 (sc-28220; Santa Cruz, CA) for the staining of pyramidal neurons. After rinsing with PBS (3 times, 5 min each), sections were stained with the corresponding secondary antibodies, followed by mounting. Cell labelling was examined with a confocal microscope (LSM510 META NLO; Carl Zeiss, Oberkochen, Germany).

**Hippocampal slice culture and transfection.** Hippocampal slices (350- $\mu$ m thick) were dissected from Sprague–Dawley rats at P7 by a vibratome (VT1200S; Leica, Wetzlar, Germany), mounted onto 0.4- $\mu$ m Millicell culture inserts (EMD Millipore, Billerica, MA). At DIV 11, slices were transfected biolistically by a PDS1000/He Biolistic Gene Gun (Bio-Rad, Hercules, CA) with 1.6- $\mu$ m gold microcarriers. At 2 to 4 days after transfection, cultures were transferred to the recording chambers and constantly perfused with oxygenated artificial cerebrospinal fluid (ACSF, 95% O<sub>2</sub> and 5% CO<sub>2</sub>) containing 125 mM NaCl, 2.5 mM KCl, 2 mM CaCl<sub>2</sub>, 1 mM MgCl<sub>2</sub>, 1.25 mM NaH<sub>2</sub>PO<sub>4</sub>, 26 mM NaHCO<sub>3</sub>, 20 mM glucose, and 200  $\mu$ M Trolox (Sigma-Aldrich, St. Louis, MO) at 29–30 °C. In some experiments, we added tetrodotoxin (Wako, Osaka, Japan, 1  $\mu$ M), bicuculline methiodide (Sigma-Aldrich, 12  $\mu$ M), lactacystin (EMD Millipore, 10  $\mu$ M) to culture and the recording medium.

**In utero electroporation.** This procedure was performed according to the published protocol with minor modifications<sup>11</sup>. Briefly, pregnant C57BL/6 mice were anaesthetized at embryonic day 13 (E13) or 14.5 (E14.5) with isoflurane, and AS-PaRac1-Venus and filler constructs (2  $\mu$ g each) were injected unilaterally into the ventricle. Electrode pulses (electrodes:  $\phi$  (diameter) 3 mm for E13 and  $\phi$  5 mm

for E14.5, 33 V, 50 ms pulse length, 950 ms pulse interval, 4 pulses) were charged unilaterally for the targeting to the M1 cortex.

**AAV viral production.** AAV viral production was performed with the AAV helper-free system (Agilent Technologies, Santa Clara, CA). The pRep-Cap (AAV5; Applied Viromics, Fremont, CA) and the pHelper plasmid were co-transfected into the AAV-293 cells with polyethylenimine 'Max' (Polysciences, Warrington, PA). After 72-h-long incubation, cells were harvested and lysed with five freeze-thaw cycles. The resultant supernatants were overlaid on 40% sucrose solution containing 100 mM Tris-HCl (pH 8.0), 150 mM NaCl, and 0.01% BSA (v/v), and were centrifuged at 100,000g for 16 h at 4 °C. The pellet (crude viral particles) was treated with 1,000 U benzonase nuclease (Novagen, Madison, WI) for 1 h at 37 °C. After filtering through a 5- $\mu$ m syringe filter to remove debris, the filtered material was subjected to CsCl gradient centrifugation (1.25 g ml<sup>-1</sup> and 1.50 g ml<sup>-1</sup>) at 257,300g for 48 h at 15 °C. The virus-rich fraction was restored, and the solvent was replaced with ASCF (1 mM MgCl<sub>2</sub>, 10 mM HEPES, CaCl<sub>2</sub>-free). Virus titre was determined with quantitative real-time PCR analysis (SYBR Green; Takara Bio Inc., Shiga, Japan).

**Virus injection and open-skull cranial window surgery.** Adult male C57BL/6 mice were anaesthetized with isoflurane, and mannitol (4  $\mu$ g per g of body weight) and dexamethasone (7  $\mu$ g per g of body weight) was administered intraperitoneally to prevent brain swelling. Subcutaneous injections of ketoprofen (40  $\mu$ g per g body weight) and penicillin/streptomycin (4 U per g body weight) were administered for 4 consecutive days beginning 1 day before the operation to prevent inflammation. The skull was exposed over the M1 cortex based on stereotaxic coordinates. Then, 1  $\mu$ l of AAV (0.5 to  $8.0 \times 10^{13}$  genome copies ml<sup>-1</sup>) was injected in the M1 cortex using a glass pipette (tip diameter 30  $\mu$ m, bevelled at an angle of 45 °) at a rate of 150 nl min<sup>-1</sup> using a syringe pump (Legato130; Muromachi Kikai, Tokyo, Japan). The location of the injection site was standardized among animals by using stereotaxic coordinates (AP = -0.8; ML = +1.0; DV = +0.5) from the skull. At the end of the injection, we waited 5 min before retracting the pipette. Stainless steel trephines ( $\phi$  2.7 mm; Fine Science Tools, Foster City, CA) were used to generate a circular open skull window. To avoid brain damage, intermittent drilling was performed at a speed of 10,000 r.p.m. with a continuous gentle perfusion of oxygenated ACSF, and we tried to avoid applying excessive drilling pressure on the skull as much as possible. If we detected no bleeding, the drilled hole was covered with a circular coverslip ( $\phi$  2.7 mm, < 0.1 mm thickness, Matsunami Glass, Kishiwada, Japan) and sealed with dental cement (Fuji Lute BC; GC, Tokyo, Japan), which was followed by the attachment of the headgear for *in vivo* imaging.

**Two-photon imaging, glutamate uncaging, and photoactivation.** Two-photon imaging was performed with an upright microscope (BX61WI; Olympus, Tokyo, Japan) equipped with an FV1000 laser scanning microscope system (FV1000, Olympus) and water-immersion objective lenses (LUMPlanFLN, 60 $\times$ , 1.0 N.A.; XLPLN25XWMP2, 25 $\times$ , 1.05 N.A.). Two mode-locked, femtosecond-pulse Ti:sapphire lasers (MaiTai DeepSee and HP; Spectra Physics, Mountain View, CA) were used at 1,000 nm for dual-colour imaging (Venus and mRFP) and at 720 nm for glutamate uncaging. For three-colour imaging of mTurquoise/GCaMP6/mRFP, the two independently captured images at 780 nm (mTurquoise and mRFP) and 970 nm (GCaMP6 and mRFP) were merged based on the identical fluorescence signal of mRFP. For *in vitro* imaging, 10–40  $\times$ y images (5 $\times$  digital zoom, 512 $\times$ 512 pixels) with a z-axis step size of 0.5  $\mu$ m were captured. For *in vivo* imaging, mice were anaesthetized with isoflurane, and images (2 $\times$  digital zoom, 1,024 $\times$ 1,024 pixels) were captured starting at the dura and progressing into the brain tissue for up to 650  $\mu$ m in total with a step size of 1.0  $\mu$ m. For glutamate uncaging, 8 mM MNI-glutamate (Tocris Bioscience, Bristol, UK) was dissolved in Mg<sup>2+</sup>-free ACSF containing 1  $\mu$ M tetrodotoxin, and using a glass pipette, this solution was applied locally onto the dendrites in the presence or absence of 10  $\mu$ M forskolin (Wako) and 5  $\mu$ M anisomycin (Sigma). Repetitive (5 Hz, 80 $\times$ ) photolysis of MNI-glutamate in the spine heads was performed at 720 nm with a pulse duration of 0.6 ms, and intensity of the uncaging laser was 6 mW under the objective lens.

**Data quantification.**  $\times$ y images were stacked by the summation of fluorescence values at each pixel. For spine size estimation, individual spines on the dendrites were traced manually, and fluorescence intensity of the filler (mRFP, DsRed Ex2, or mTurquoise) was measured in the spine-head. For each channel, background intensity was subtracted from the fluorescence intensity (arbitrary units, a.u.) of each spine. During time-lapse imaging, daily variations in the recording conditions caused slight alterations in the fluorescence intensity, which was corrected with the fluorescence intensity changes of the filler along the parental dendritic shaft within a distance of 10  $\mu$ m from the spine. The 'Spine enrichment index' was estimated based on the previous report<sup>20</sup>. To assess the uneven distribution of



PaRac1 variants in the dendrite, the 'Hot spot index' was calculated using the following equations:

$$\frac{1}{n} \times \sum_{i=1}^n |(\text{Spine enrichment index}_i - \text{Spine enrichment index}_{i+1})|$$

where 'Spine enrichment index<sub>*i*</sub>' and 'Spine enrichment index<sub>*i+1*</sub>' represent the enrichment indices of a given spine and of its nearest neighbouring spine, respectively, and 'n' represents the number of spines in the measured dendritic branch (20 µm long). Hot spot index was obtained from the most intensively labelled dendritic segments, and estimated by repetitive measurements of sequential nearest neighbouring spines. Quantification of fluorescence was performed with the ImageJ software.

**In vivo photoactivation in freely moving animals.** Mice transduced by either AAV injection or *in utero* gene transfer were subjected to open-skull cranial window surgery, and the cranial holes were covered with bilateral glass windows. An outer cylinder (a non-bevelled 15-mm long 18G needle with an inner diameter of 0.9 mm) was implanted on the glass window for photoactivation. Before photoactivation, the optical fibre was inserted into the outer cylinder, and the tip of the fibre was placed directly onto the glass coverslip. The fibre and the outer cylinder were tightly locked together with Blu-Tack, which was easily removed after the photoactivation. Photostimulation was carried out using the COME-2 series (Lucir, Osaka, Japan), which consist of 457-nm laser diodes, an optical swivel, and bilateral optical fibres (COME2-αDF1; core diameter of 500 µm, 0.5 N.A.). The laser diode was adjusted to an output of 20 mW at the tip of each fibre. The light pulse was delivered for 150 ms at 1 Hz for 1 h, and the process was controlled by customised LabView programs (National Instruments, Austin, TX).

**Behavioural analysis.** Mice were housed under standard laboratory conditions (12-h light/dark cycle with food and water available *ad libitum*) and were randomly allocated to experimental groups. All behavioural analyses were performed during the light phase. For motor learning (Extended Data Fig. 9), we used the rotarod training system (Rota-Rod Treadmills ENV-576; Med Associates, St. Albans, VT). Before the training sessions, mice were habituated to stay on the stationary rod for 2 min. During the training period, the fixed-speed protocol was applied at a slow speed (8 rpm), so mice rarely fell off the rod. After the mice were able to remain on the rod reliably, the speed was increased in a stepwise fashion to 40 rpm. We applied air puffs to the hind limbs as aversive stimuli to teach mice to face forward on the rod (Extended Data Fig. 9c), which helped them to hold on at higher speeds. After falling, the mice were immediately placed back on the rod, and latency of falling was recorded automatically. Three training sessions were performed for 2 days (2 h for 1 session, 6 h of training in total). To assess learning, three trials of the rotarod test were carried out using an accelerating protocol (4 to 40 rpm) without air puffs with 5 min inter-trial intervals.

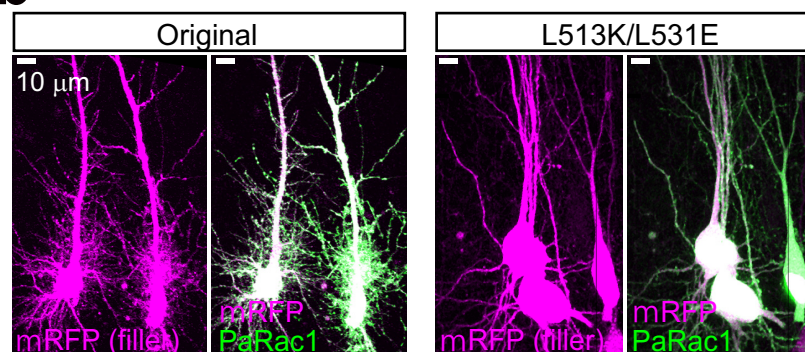
For balance beam training, a hand-made beam apparatus was used (Extended Data Fig. 9d). Time to cross was scored using a stopwatch. The timer was started when the mouse was placed on the beam and ends when the first forepaw was placed in the goal cage. Air puffs to the hind limbs were also used to facilitate learning. Three training sessions were performed during 2 days (2 h for 1 session, 6 h of training in total). To evaluate the acquired performance, three trials of the beam test were carried out without air puffs with 5 min inter-trial intervals. Task performances were calculated as the averages of the three trials for both the rotarod and beam tasks. Mice with an improvement of < 20% compared to the pre-training performance were excluded from the analyses. The running speed of mice was measured by a video tracking system (Limelight3; Actimetrics, Wilmette, IL). The investigator was not blinded to the group allocation during the experiments because all behavioural outcomes were unambiguously determined: for example, rotarod performance and locomotion were scored automatically with infrared or video tracking, and the manual scoring of the cross time for the beam test was unambiguous.

**Statistics.** A series of experiments were performed as two, mostly three separate cohorts, and sample size was chosen based on the effect size shown in the first cohort in order to minimize the number of animals used in compliance with ethical guidelines. Data are shown as means ± s.e.m. Detailed information on statistical methods/results are described in Extended Data Table 1. In brief, Mann-Whitney U tests were used to identify significant differences between two groups. Multiple comparisons were made by one-way analysis of variance (ANOVA, normal distribution and equal variances), nonparametric one-way ANOVA (Kruskal-Wallis test, for unequal variances), or one-way repeated measures ANOVA followed by post-hoc Bonferroni test (to compare task performance at different time points for within-subjects groups). Spearman rank correlation was used to test the strength of correlation between two variables. For all statistical tests \**P* < 0.05, \*\**P* < 0.01, \*\*\**P* < 0.001 were considered significant. No statistical methods were used to predetermine sample size, the experiments were randomized and the investigators were not blinded to outcome assessment.

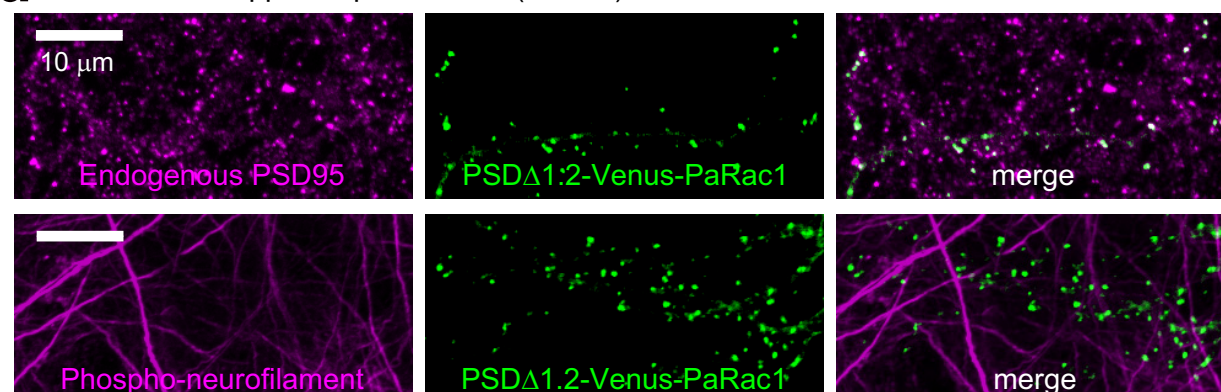
44. Andreassi, C. & Riccio, A. To localize or not to localize: mRNA fate is in 3'UTR ends. *Trends Cell Biol.* **19**, 465–474 (2009).
45. Wang, D. O., Martin, K. C. & Zukin, R. S. Spatially restricting gene expression by local translation at synapses. *Trends Neurosci.* **33**, 173–182 (2010).
46. Holt, C. E. & Schuman, E. M. The central dogma decentralized: new perspectives on RNA function and local translation in neurons. *Neuron* **80**, 648–657 (2013).
47. Gray, N. W., Weimer, R. M., Bureau, I. & Svoboda, K. Rapid redistribution of synaptic PSD-95 in the neocortex *in vivo*. *PLoS Biol.* **4**, e370 (2006).
48. Bosch, M. *et al.* Structural and molecular remodeling of dendritic spine substructures during long-term potentiation. *Neuron* **82**, 444–459 (2014).
49. Hsueh, Y. P., Kim, E. & Sheng, M. Disulfide-linked head-to-head multimerization in the mechanism of ion channel clustering by PSD-95. *Neuron* **18**, 803–814 (1997).
50. Kasai, H. *et al.* Learning rules and persistence of dendritic spines. *Eur. J. Neurosci.* **32**, 241–249 (2010).

**a**

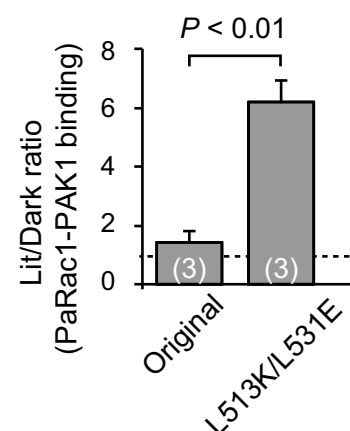
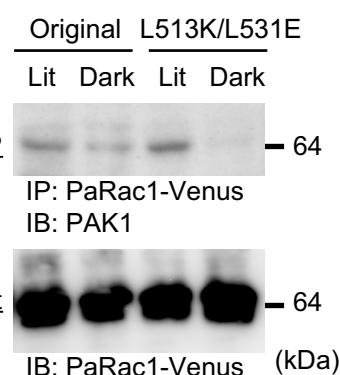
Experiment	$K_d$ ( $\mu$ M)	# of trials
PaRac1 (C450A, light-insensitive mutant)	$2.21 \pm 2.14$	7
PaRac1 (I539E, Lit state mutant)	$0.197 \pm 0.039$	8
PaRac1 (Original), Dark	$3.18 \pm 0.57$	4
PaRac1 (Original), Lit	$0.193 \pm 0.014$	4
PaRac1 (L514K/L531E), Dark	$5.52 \pm 1.17$	4
PaRac1 (L514K/L531E), Lit	$0.362 \pm 0.051$	13

**b****c**

Dissociated hippocampal neurons (DIV 21)

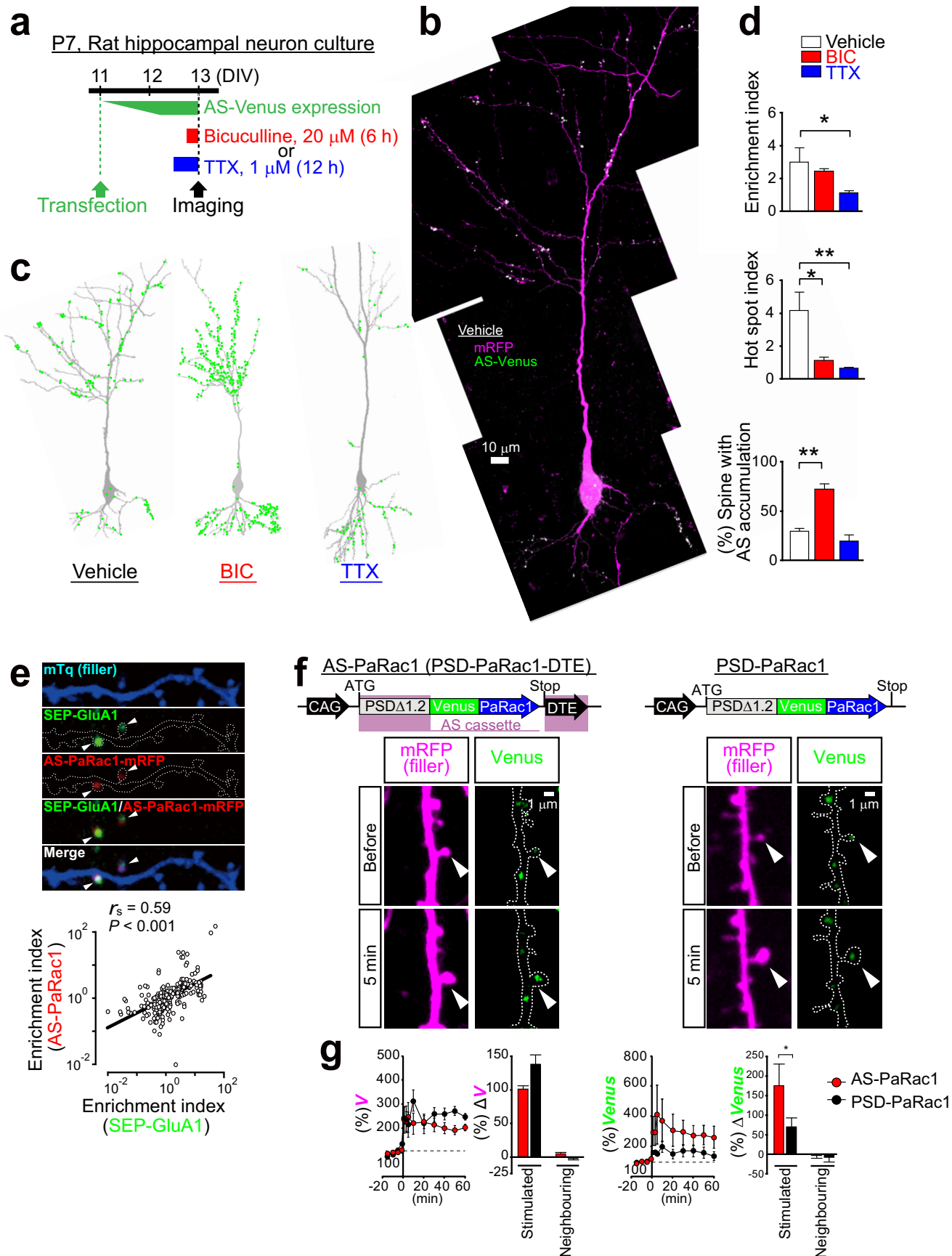


**Extended Data Figure 1 | Optimization of the PaRac1 for the synaptic application.** **a**, Isothermal titration calorimetry (ITC) experiments showing that the introduction of L514K and L531E mutations into the original PaRac1 construct<sup>12</sup> reduced binding with the CRIB domain of PAK1 in the dark. The light-insensitive form of LOV2(C450A) and the I539E mutant, which mimics the unfolded 'lit state', were used as negative and positive controls, respectively. **b**, Leaky activity of PaRac1 in the dark. In hippocampal neuronal cultures transfected with the original PaRac1, we observed a bearded appearance of the soma with numerous ectopic dendrites, while neurons transfected with PaRac1 (L514K/L531E) were indistinguishable from normal neurons. **c**, Assessment of the affinity of PaRac1 to the endogenous PAK1 using a pull-down assay. HEK293 cells, which were transfected with PaRac1-Venus, were divided into two groups: lit and dark. The cells in the lit group were radiated with light with a white fluorescent lamp before cell lysis, and

**c**

continuous light illumination was present during subsequent immunoprecipitation until the final wash step of protein precipitants. Conversely, cells in the dark group were lit with a yellow fluorescence lamp, which excludes light wavelengths below 500 nm. Co-immunoprecipitation with PAK1 revealed that PaRac1 (L514K/L531E) barely bound with PAK1 in the dark (the number of trials is depicted in the bar graph,  $^{**}P < 0.01$  using the Mann-Whitney U test). **d**, Targeting of PaRac1 to the postsynaptic density. PSDΔ1.2-PaRac1 (DTE (-)) was transfected into dissociated cortical neurons at 21 days *in vitro* (DIV). Two days after transfection, cells were fixed with 4% PFA, followed by permeabilization for the subsequent immunostaining procedure. Axons and endogenous PSD-95 were visualized using the anti-phospho-neurofilament and anti-PSD-95 antibodies, respectively, revealing that PSDΔ1.2-PaRac1 co-localized with the endogenous PSD-95. Note that PSDΔ1.2-PaRac1 did not co-localize with the axonal marker.

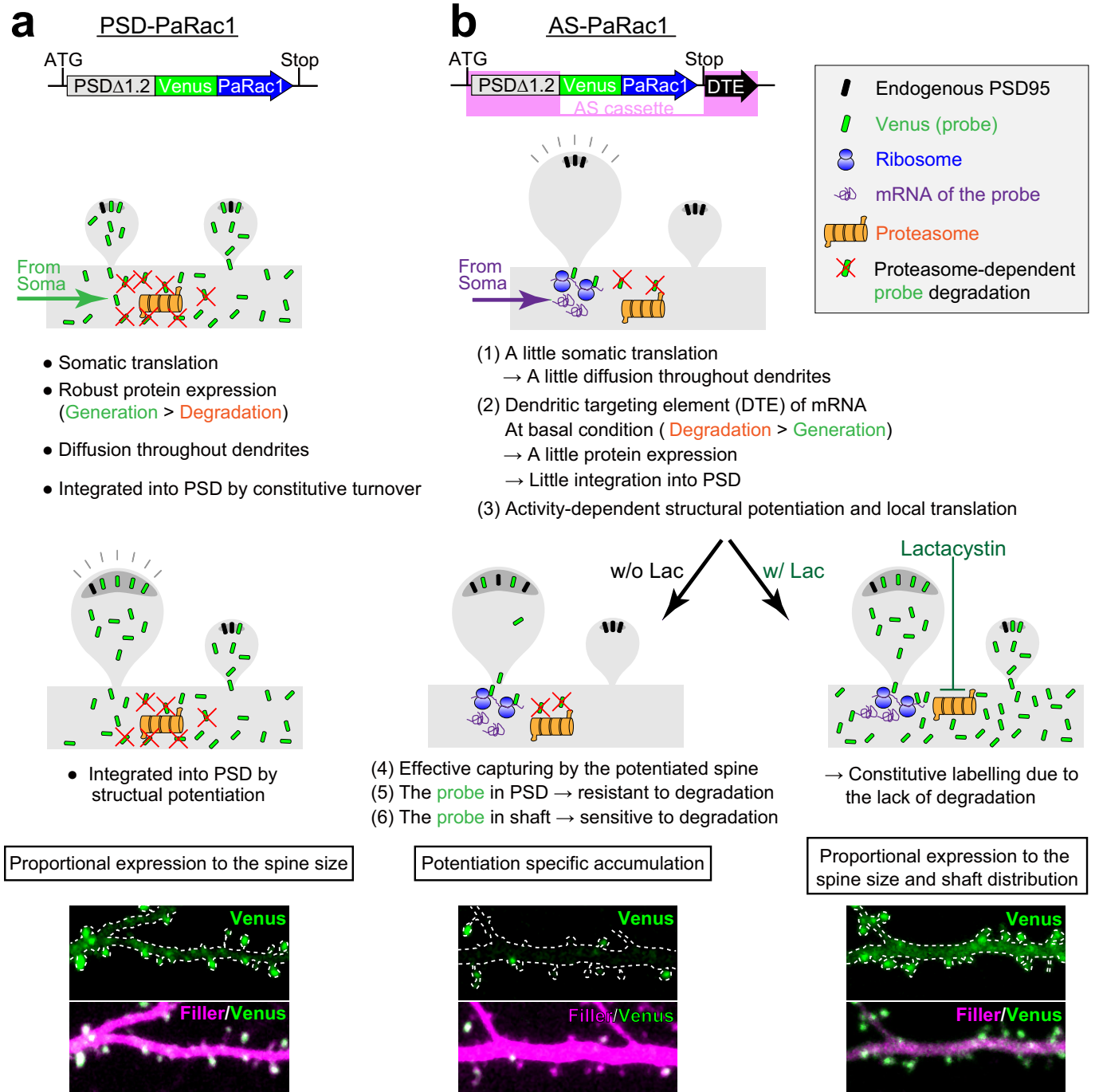




**Extended Data Figure 2 | The distribution of AS-PaRac1 is regulated by neuronal activity, and is dependent on the dendritic targeting element (DTE).** **a**, Experimental design. **b**, Representative image of a cultured hippocampal neuron. **c**, Bicuculline (BIC) or tetrodotoxin (TTX) was added to the culture media at the designated time points. Images were captured at a high magnification and were tiled to visualize the entire cell. Green circles represent the AS-PaRac1 puncta. **d**, Quantification of AS-PaRac1 distribution ( $n = 6$  each,  $*P < 0.05$ ,  $**P < 0.01$  using Kruskal–Wallis test followed by post-hoc Dennett's test). **e**, Concomitant accumulation of AS-PaRac1 and SEP–GluA1 in spines. Neurons were co-transfected with mTq (mTurquoise, filler), SEP–GluA1, and AS-PaRac1–mRFP, and the constructs were expressed for 36 h. Potentiated spines during 36 h were shown by SEP–GluA1 fluorescence (arrowheads). Spearman rank correlation revealed a significant correlation between the spine enrichment indices of SEP–GluA1 and AS-PaRac1 (each

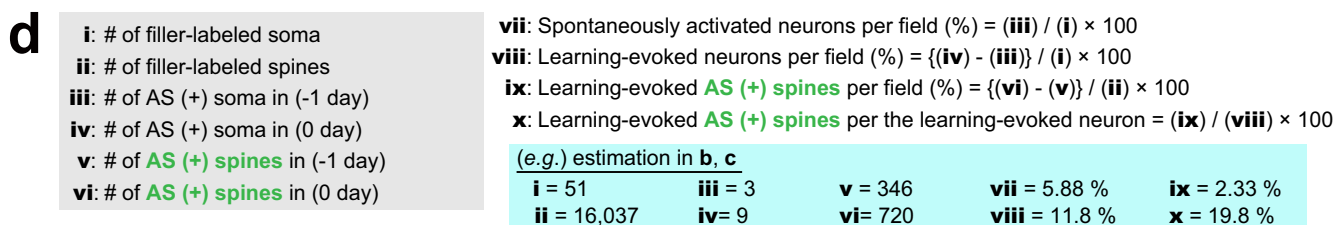
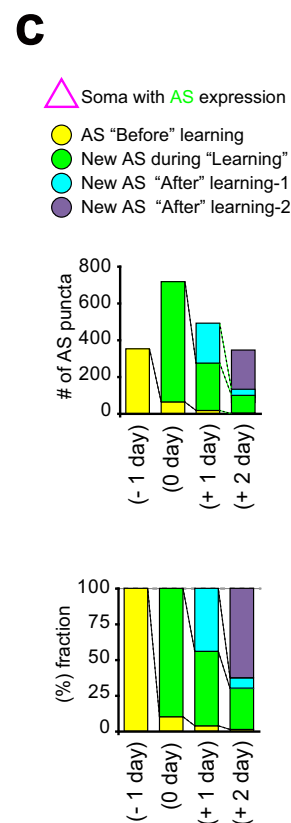
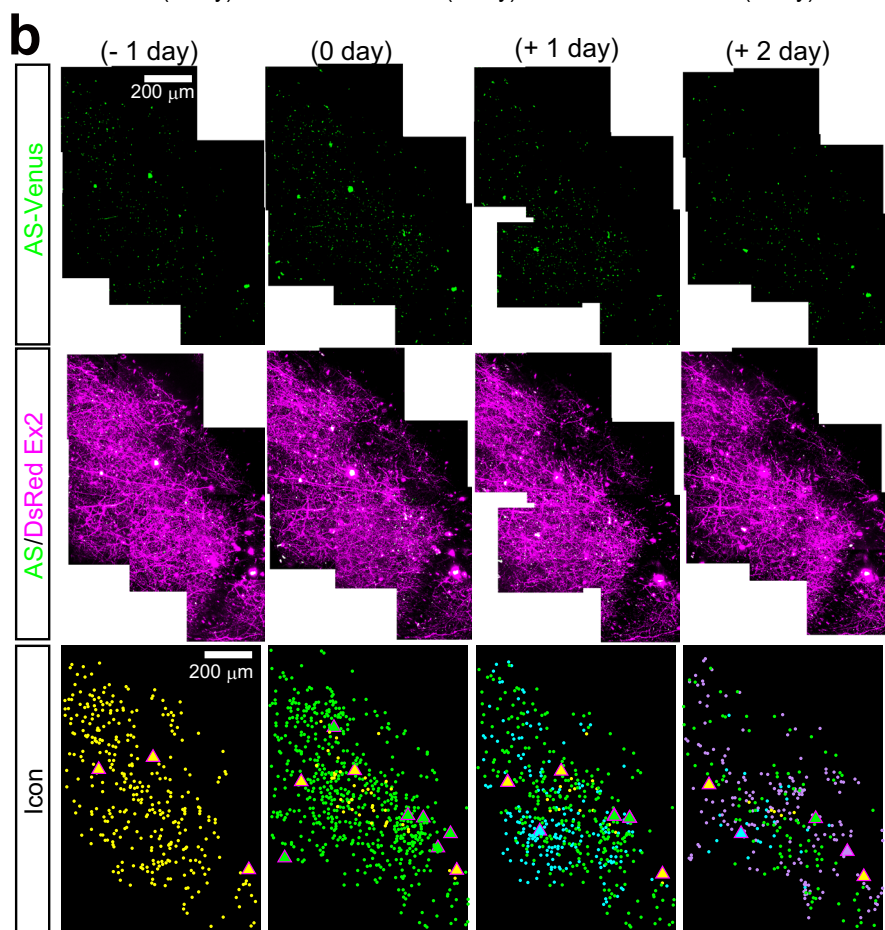
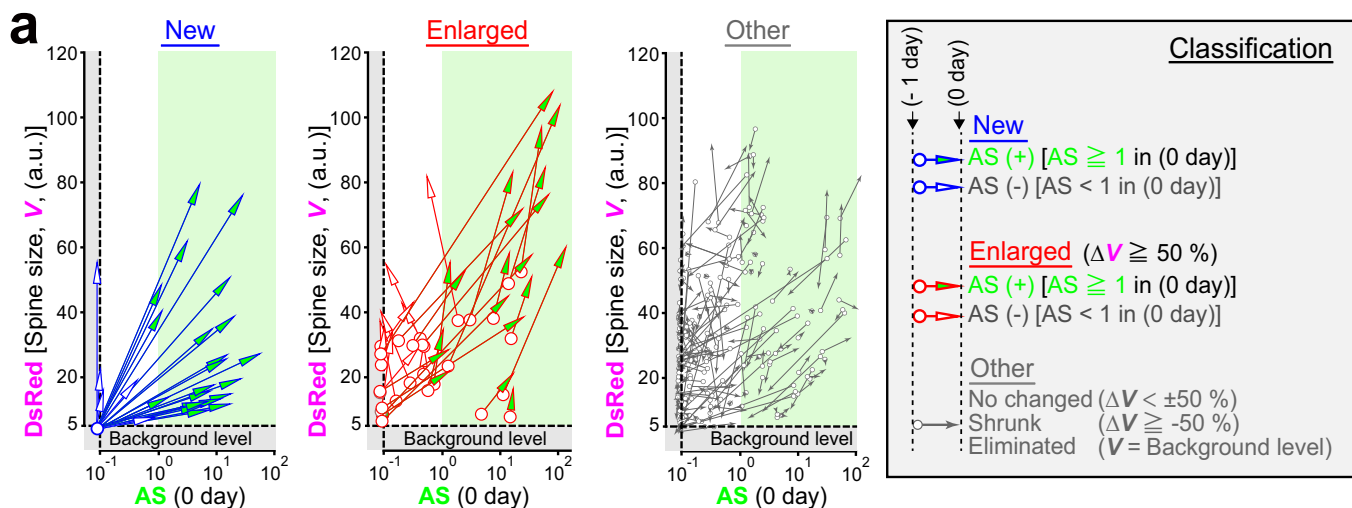
circle represents one spine, 235 spines, 29 dendrites). **f**, Schematic of the constructs and representative images of single spine potentiations by glutamate uncaging in the presence of FSK (arrowheads). Rat hippocampal slice cultures were biolistically transfected with either AS-PaRac1 or PSD–PaRac1 (DTE (–)) followed by the uncaging experiments at DIV 13 (equivalent to postnatal day 20). **g**, Time course of the spine head volume ( $V$ ) and accumulation of Venus upon uncaging. The mean changes in spine size and Venus accumulation in the stimulated or neighbouring spines are depicted 60 min after uncaging. For quantification, we used pooled data from independent identically designed experiments. The data set for AS-PaRac1 was identical with the FSK-treated group of Fig. 1c–e. Scale bars, 1  $\mu\text{m}$ .  $*P < 0.05$  using the Mann–Whitney U test ( $n = 6$  or 11 dendrites for PSD–PaRac1 or AS-PaRac1, respectively).





**Extended Data Figure 3 | Putative cellular mechanisms of the specific concentration of AS-PaRac1 in potentiated spines.** **a**, Uniform labelling of spines with the PSD–PaRac1 construct that lacks DTE of *Arc* 3' UTR (Fig. 1a, construct B). PSD–PaRac1 is translated in the soma that is abundantly equipped with translational machineries. Therefore, the somatic protein expression of the probe is high (data not shown), which would outnumber the degradation, and the resulting proteins are transported throughout dendrites. The overflowing probes integrate into the postsynaptic density (PSD) during the constitutive turnover of PSD molecules. Therefore, probe expression is proportional to the spine size. **b**, Selective labelling of potentiated spines with AS-PaRac1 (Fig. 1a, construct C). The following six mechanisms endow the potentiation-specific labelling with AS-PaRac1. (1) A little somatic translation: the moderate gene expression of AS-PaRac1, by which the translation of AS-PaRac1 protein is limited in the soma (see Extended Data Fig. 2b), and therefore, the non-specific overflow of this probe from the soma into the dendrites is minimal. (2) Dendritic targeting element (DTE): the essential domains of AS-PaRac1 are the N-terminal PSD-95 (PSDΔ1.2) and the 3' UTR of *Arc* mRNA (DTE). DTE has a pivotal role in the dendritic targeting of mRNAs<sup>44,45</sup>. One of the most well-known DTE is present in the *Arc* mRNA<sup>16</sup>, which is targeted to stimulated dendritic segments in an activity-dependent manner<sup>18</sup>. The transport of mRNA out of soma also contributes to the limited translation of the probe in the soma described in (1). In the absence of activation, the limited amount of translational machineries and presence of

degradation components in the dendrites maintains the locally translated probe at a low level, which results in a low rate of AS-PaRac1 integration into the PSD during the constitutive turnover of PSD proteins. (3) Local protein synthesis: persistent structural plasticity of the spine depends on the activity-dependent dendritic synthesis of proteins<sup>46</sup>, and the translation of *Arc* mRNA is controlled by activity levels<sup>19</sup>. (4) Effective capturing of PSD proteins in the structurally potentiated spines: the potentiated spine, which rapidly requires new copies of PSD proteins, captures diffusing PSD proteins more efficiently<sup>47,48</sup>. (5) Increased stability of AS-PaRac1 in the PSD: it is likely that the stability of the PSD-integrated AS-PaRac1 increase, as does the typical PSD scaffold proteins<sup>47</sup>. The ubiquitination might be underlying mechanism of the increased stability, because the ubiquitination site of AS-PaRac1 resides in the N-terminal domain of PSD-95, the domain of which is aggregated to form head-to-head multimerization in the postsynaptic scaffold<sup>49</sup>. Thus, once AS-PaRac1 is integrated into the PSD, the ubiquitination site may be concealed, and AS-PaRac1 becomes relatively stable. (6) Sensitivity of unbound AS-PaRac1 against the proteasomal degradation: contrary to the PSD-integrated AS-PaRac1, unbound AS-PaRac1 is sensitive to degradation because the ubiquitination site is not concealed. This scenario is supported by the administration of lactacystin (right panel), which inhibits proteasomes and thus completely disrupts the uneven distribution of AS-PaRac1. Similar mechanisms are relevant for newly formed spines, because spine formation is associated with spine enlargement<sup>50</sup>.



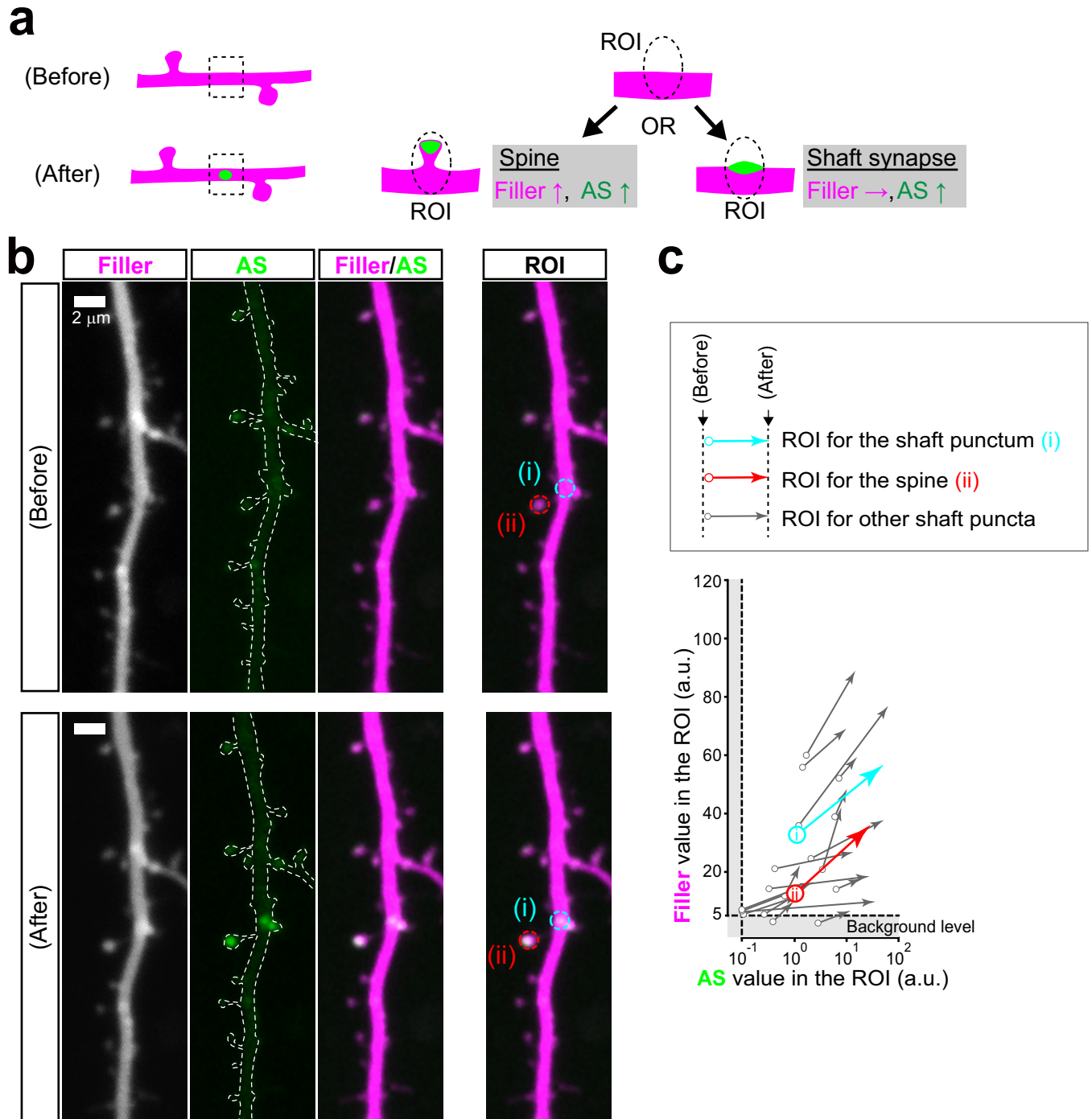
	viii: Learning neuron (%)	ix: Learning spines (%)	x: Learning spines / Learning neurons (%)	# of mice
Layer II/III neuron	16.4 ± 2.8	2.3 ± 0.13	14.71 ± 2.01	5
Layer V neuron	22.57 ± 2.8	1.15 ± 0.27	5.01 ± 0.76	4



**Extended Data Figure 4 | Raw data of quantification and synaptic mapping.**

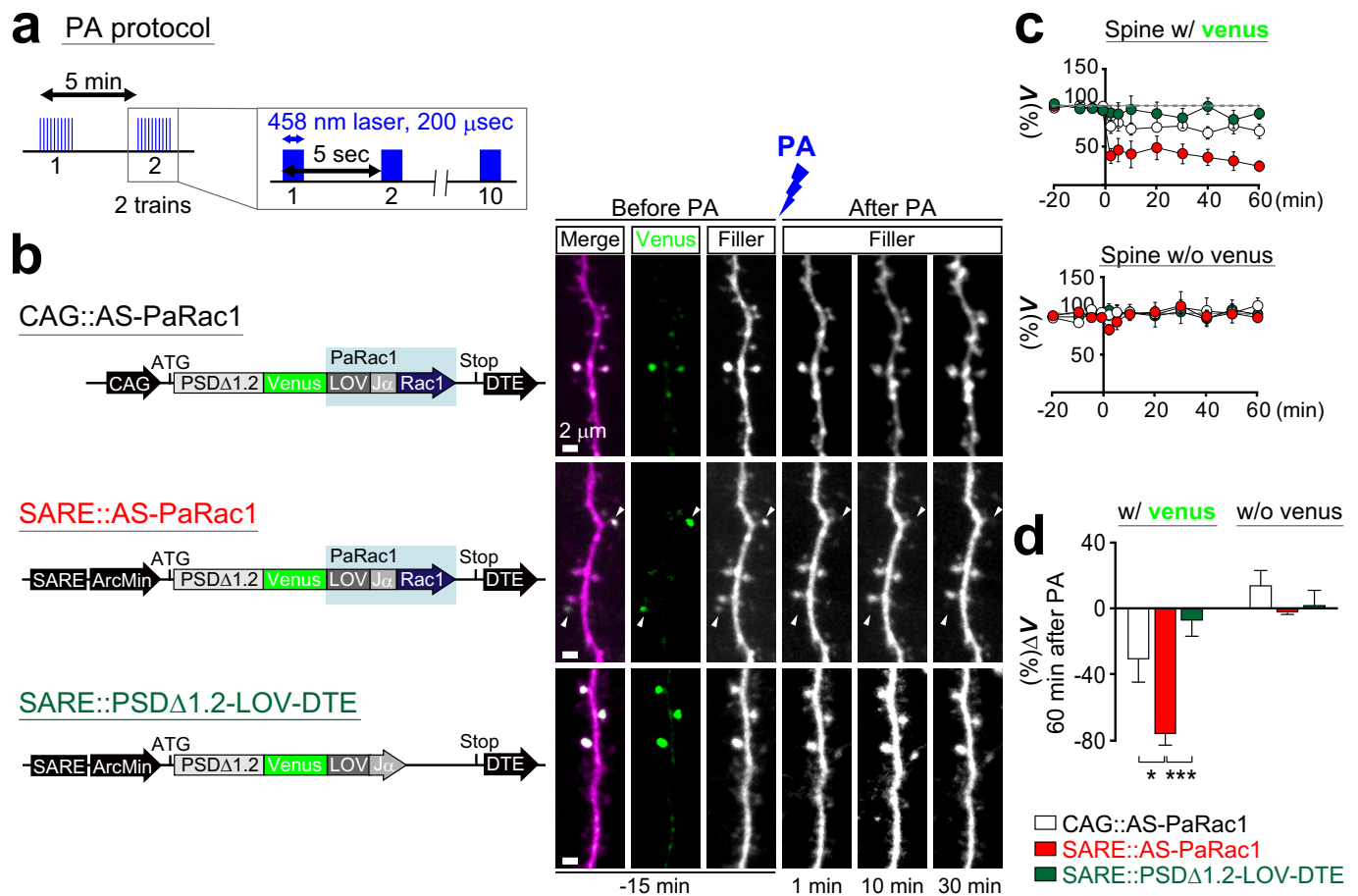
Data from Fig. 2. **a**, Quantification of spine size (based on DsRed fluorescence) and AS-PaRac1 fluorescence after learning are depicted separately based on the classification of spines. The definitions of 'New spine', 'Enlarged spine' and others are described on the right. Each arrow indicates the trajectory of a spine; beginning and end points represent the absolute values before and after the rotarod task, respectively. **b**, *xy* images were captured from the dura to a depth of 300  $\mu\text{m}$  with a step-size of 1.0  $\mu\text{m}$ , and were stacked by the summation of fluorescence values at each pixel. *z*-stacked images of 10 overlapping fields were aligned to generate the combined images. AS-PaRac1 and AS-PaRac1/

DsRed merged images are shown. AS-PaRac1 that was present before learning ( $-1$  day, yellow), appeared shortly after learning (learning period, 0 day, green), 1 day (after-1,  $+1$  day, blue), or 2 days after learning (after-2,  $+2$  day, purple) are depicted to show the spatiotemporal distribution of AS-PaRac1 triggered in each period. **c**, Time course of the number and fraction of AS-PaRac1-positive spines in each period. **d**, Calculation of the learning-evoked spine/neuron ratio (%). Example of the calculation is based on the raw data shown in **b** and **c**. The table indicates the comparison between neurons in layer II/III (*in utero* electroporation at E14.5) and layer V (*in utero* electroporation at E13).



**Extended Data Figure 5 | Assessment of AS-PaRac1 puncta on the dendritic shaft.** **a**, The two possible synapse types that AS-PaRac1 puncta may represent on the dendritic shaft. xy images were captured to encompass the entire z-range of the dendrite of interest with a step-size of 0.5  $\mu\text{m}$ , and images were stacked by the summation of fluorescence values at each pixel. The fluorescence of both the filler and AS-PaRac1 would increase, if the AS-PaRac1 punctum emerged on the dendritic spine that undergoes structural potentiation. In contrast, fluorescence of the filler would not increase, if AS-PaRac1 was in the shaft synapse. **b**, Example of the dendrites before and after the emergence of AS-PaRac1. AS-PaRac1 puncta on the shaft and on the dendritic spine are indicated with (i) and (ii), respectively. The region of interest (ROI) used for

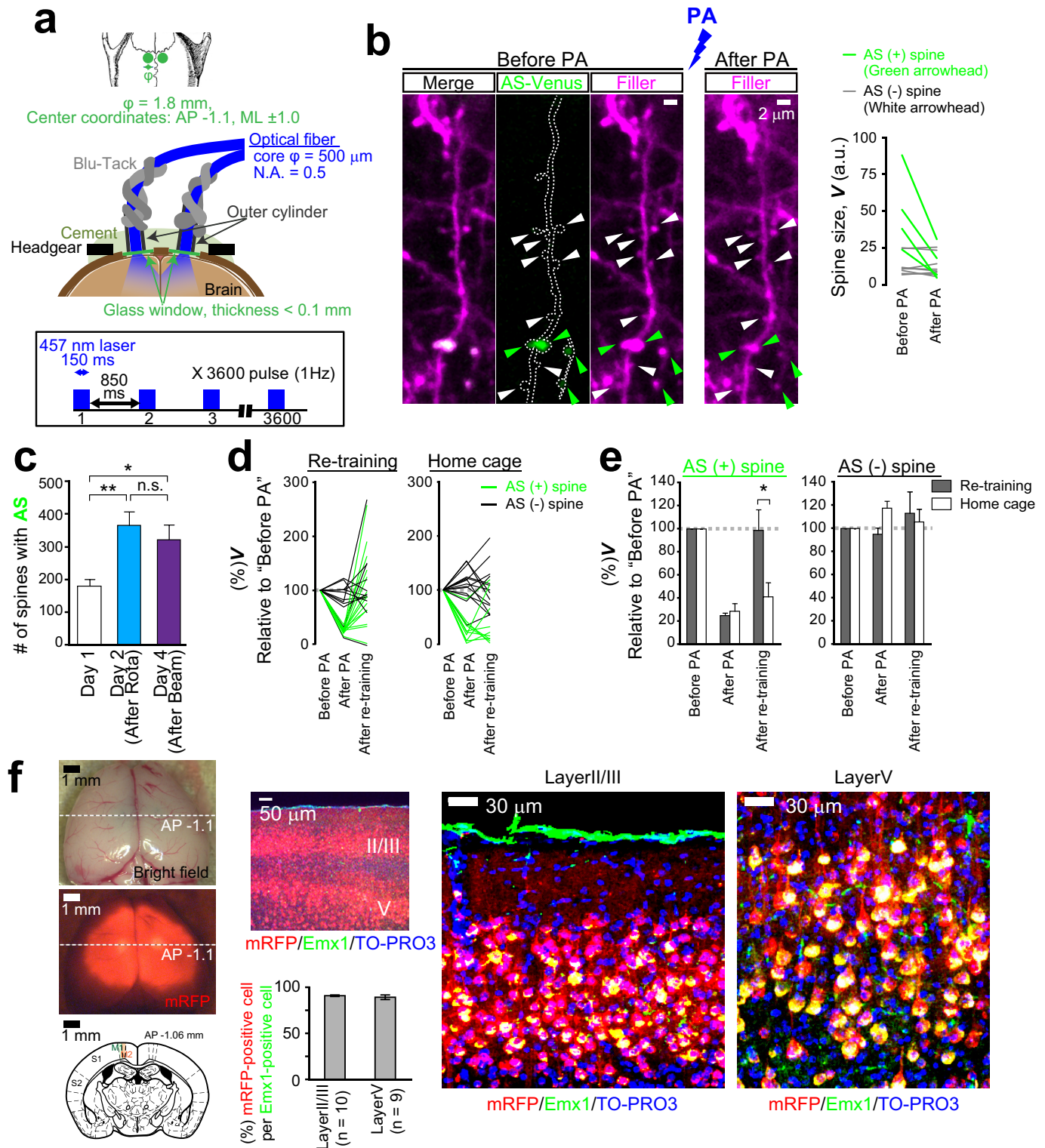
the calculation of fluorescence in each punctum is shown. **c**, Quantification of the fluorescence of the filler and AS-PaRac1 upon the emergence of AS-PaRac1 puncta. Each arrow indicates the trajectory of each ROI; beginning and end points represent the absolute values before and after the emergence of AS-PaRac1, respectively. The ROI at (i) exhibited a concomitant fluorescence increase in both the filler and AS-PaRac1, similar to AS-PaRac1 in a typical dendritic spine (ii). All examined AS-PaRac1 puncta on the dendritic shaft exhibited positive correlations, suggesting that the majority of AS-PaRac1 puncta emerge on the dendritic spine during the structural changes of the spine.



**Extended Data Figure 6 | Rac1-dependent shrinkage of dendritic spines induced by low-frequency photoactivation.** **a**, The protocol of photoactivation. Photoactivation was performed in the region that encompasses the branch of interest. **b**, Neurons in the hippocampal slice culture (DIV 11) were biologically transfected with DNA constructs shown in the schematic image on the left. Representative dendritic images upon photoactivation are shown on the right. Robust shrinkage (arrowheads) was observed in the spines transfected with AS-PaRac1 driven by the SARE-Arc promoter. Despite their adjacent location to the AS-PaRac1-positive spines,

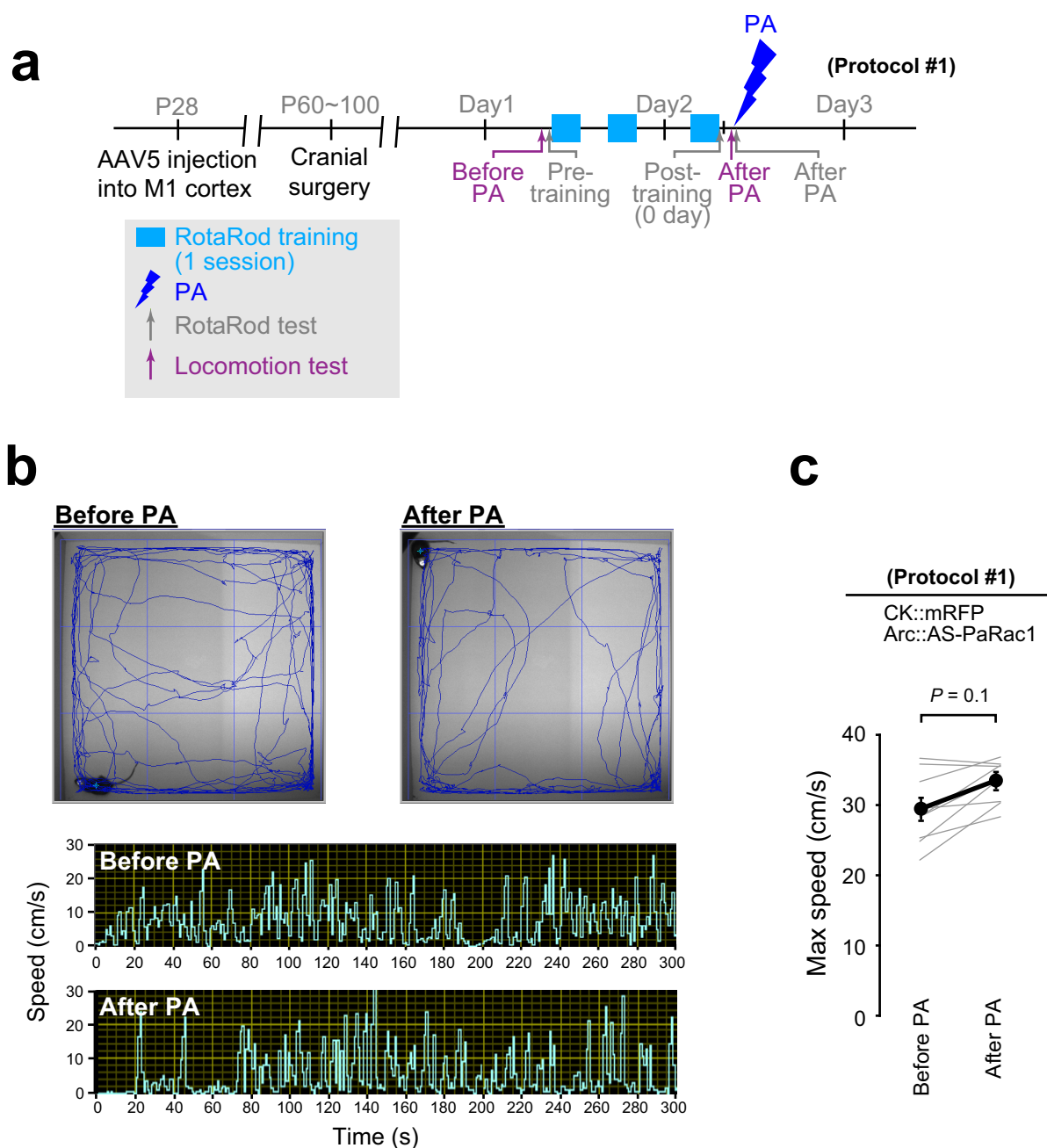
AS-PaRac1-negative spines were not affected by the photoactivation. **c**, Time course of the spine head volume ( $V$ ) of Venus-positive (upper panel) and negative spines (lower panel). White, red, and blue circles represent CAG::AS-PaRac1, SARE::AS-PaRac1, and SARE::PSDΔ1.2-LOV-DTE, respectively ( $n = 12$  cells each). **d**, The mean relative change in spine head size in Venus-positive and negative spines 60 min after photoactivation. Scale bars, 2  $\mu\text{m}$ .  $*P < 0.05$  and  $***P < 0.001$  according to the Kruskal–Wallis test followed by the post-hoc Scheffé's test.





**Extended Data Figure 7 | Spine shrinkage in broad areas of the bilateral motor cortices induced by blue laser illumination.** **a**, Schematic of the bilateral cranial windows, optical fibres, and the photoactivation protocol. **b**, Representative images of spine shrinkage in the M1 cortex upon photoactivation *in vivo*. AS-PaRac1-positive spines (green arrowheads) shrank, while the AS-PaRac1-negative ones (white arrowheads) did not. Quantification of spine size is shown on the right. **c**, The mean number of AS-PaRac1 puncta per fields was calculated in mice shown in the Fig. 4i. **d**, **e**, Spine structure and AS-PaRac1 were imaged in mice, which were subjected to the

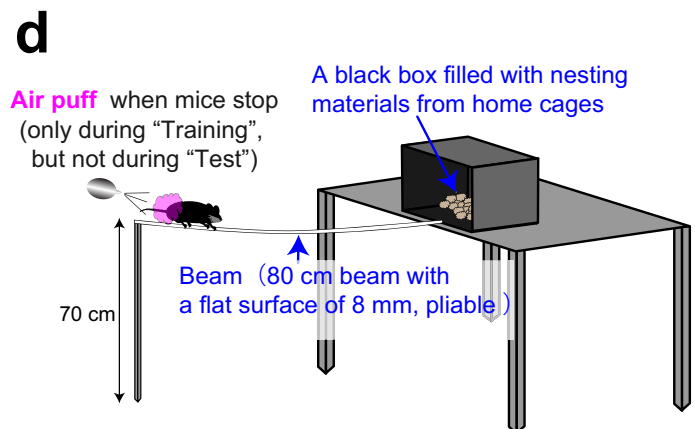
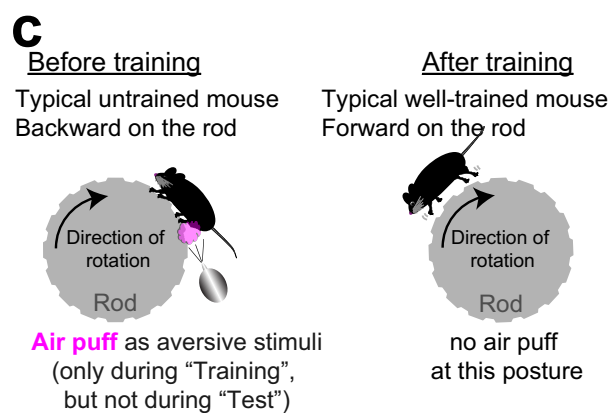
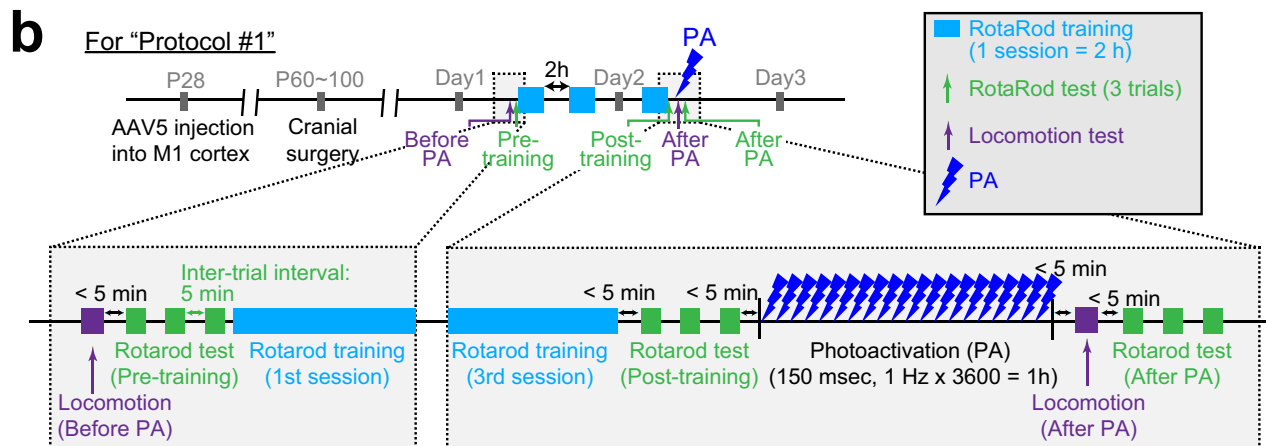
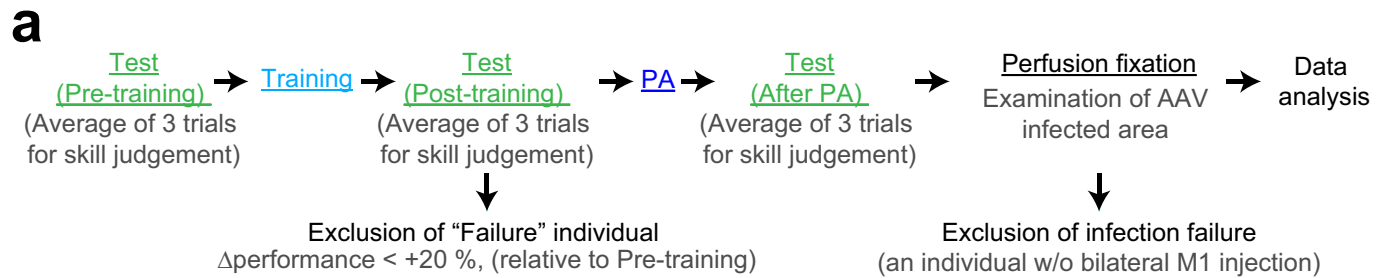
re-training and home cage protocols shown in Fig. 5. The majority of AS-PaRac1-positive spines displayed photoactivation-induced shrinkage and subsequent recovery.  $*P < 0.05$  according to the Mann-Whitney U test. **f**, The success of AAV5 vector injection into the bilateral M1 cortex was confirmed by the presence of mRFP fluorescence after behavioural tests. High efficacy of virus infection in layer II/III and V pyramidal neurons was demonstrated with Emx1 immunostaining, which labels pyramidal neurons. The mice without bilateral mRFP signal in the M1 cortex were excluded from the data analysis.



**Extended Data Figure 8 | No effect of photoactivation on the locomotor activity of mice.** **a**, Experimental schedule. The running speed of AS-PaRac1-injected mice in protocol no. 1 (Fig. 4a) was measured with a video-tracking system. To minimize the effect of circadian rhythm on locomotion, mice were

tested at the same time of the day before and after photoactivation.

**b**, Representative traces of locomotion and temporal sequences of running speed are depicted. **c**, Statistical analysis shows that photoactivation has only a negligible effect on running speed.



**Extended Data Figure 9 | Detailed illustration of the rotarod and beam tasks.** Experimental setup for Fig. 4. **a**, Experimental flowchart. **b**, Detailed schedule of the rotarod training/test, locomotion test, and photoactivation. **c**, To shorten the training time, air puffs were applied to the hind limbs as

aversive stimuli to maintain the forward-looking position of mice on the rod, which improved the performance, especially at higher speeds. **d**, Schematic illustration of the beam test. The test was preceded by a 6-h-long training session that lasted for 2 days.



Extended Data Table 1 | Detailed information on sample descriptions and statistics

	Sample			Statics												
	Description	Size (n)	Methods	Comparison	P values				Correlation coefficient							
Figure 1a	Hippocampal slice culture + Gene Gun	Construct (A) = 13 dendrites/13 slices/3 rats	One-way factorial ANOVA (post-hoc Dunnet test)	(A) vs (B) (A) vs (C) (A) vs (D) (A) vs (E)	Enrichment	Hot spot										
		Construct (B) = 20 dendrites/20 slices/6 rats			0.03896	0.48912										
		Construct (C) = 23 dendrites/23 slices/6 rats			0.00001	0.00000										
		Construct (D) = 8 dendrites/8 slices/3 rats			0.69713	0.81024										
		Construct (E) = 8 dendrites/8 slices/3 rats			0.00130	0.00929										
Figure 1b-d			Kruskal Wallis test (post-hoc Scheffe's test)	(A) vs (B) (A) vs (C) (B) vs (C)	mRFP	Venus										
		uncaging alone (A) = 15 dendrites/15 slices/6 rats			0.62258	0.01139										
		uncaging + FSK (B) = 35 dendrites/35 slices/8 rats			0.00430	0.10843										
Figure 1e		uncaging + Aniso (C) = 20 dendrites/15 slices/6 rats			0.00000	0.00000										
			Enrichment	Hot spot												
		Vehicle = 13 dendrites/13 slices/3 rats	Mann-Whitney test (two-sided)	Veh vs Lac	0.00134	0.00030										
		Lactacystin = 8 dendrites/8 slices/3 rats														
Figure 2d	In vivo M1 cortex + In utero EP (E14.5)	Training = 2793 spines/7 mice	Mann-Whitney test (two-sided)	(Training) vs (No training)	Enlarged	New	Shrunk	Eliminated								
No training = 718 spines/3 mice		0.03887			0.02014	0.12134	0.12134									
Figure 2f		Training = 2090 spines/3 mice	Spearman's rank correlation coefficient	$\Delta V$ (0 day) & AS (0 day) $\Delta V$ (0 day) & AS (-1 day)	0.00000	0.51746			0.61278							
Figure 2g									-0.03549							
Figure 2k			68 spines (enlarged or new spines) out of 2090 total spines for Fig 2e-j	Mann-Whitney test (two-sided)	[AS (+1 day) $\geq$ 1] vs [AS (+1 day) < 1]	0.00263	0.00016									
Figure 3c	In vivo M1 cortex + In utero EP	94 spines/6 mice	Mann-Whitney test (two-sided)	[AS (+)] vs [AS (-)]	0.00000											
Figure 3d			Spearman's rank correlation coefficient	$\Delta V$ & Depth	0.28151			-0.17225								
Figure 3i, j	Hippocampal slice culture + Gene Gun	24 spines/12 slices/6 mice	Spearman's rank correlation coefficient	$\Delta V$ & $\Delta$ Amp $\Delta V$ & $\Delta$ Freq	AS (+)	AS (-)			AS (+)	AS (-)						
					0.01793	0.57016			0.58235	0.23810						
Figure 3i, j		16 soma/12 slices/6 mice	Wilcoxon signed rank test	(Before PA) vs (After PA)	Amplitude	Frequency			-0.14706	0.02381						
					0.24886	0.09620										
Figure 4b-d	In vivo M1 cortex + Bilateral AAV5 infection	mRFP alone (Protocol #1) = 10 mice mRFP + AS-PaRac1 (Protocol #1) = 15 mice mRFP + AS-PaRac1 (Protocol #2) = 12 mice mRFP + AS-PaRac1 (Protocol #3) = 5 mice	One-way repeated measures ANOVA (post-hoc Bonferroni test)	(Pre-training) vs (0 day) (0 day) vs (+1 day) (0 day) vs (+2 day) (0 day) vs (After PA) (+1 day) vs (After PA) (+2 day) vs (After PA) (Pre-training) vs (After PA)	mRFP (Prot #1)	AS (Prot #1)	AS (Prot #2)	AS (Prot #3)								
					0.00090	0.00000	0.00000	0.00034								
					NA	NA	0.83880	NA								
					NA	NA	NA	1.00000								
					0.59421	0.00056	0.00003	1.00000								
					NA	NA	0.00000	NA								
					NA	NA	NA	1.00000								
					0.00005	0.00070	0.01852	0.00012								
					Figure 4e		One-way factorial ANOVA (post-hoc Scheffe's test)	(Prot #1) vs (Prot #2) (Prot #2) vs (Prot #3) (Prot #1) vs (Prot #3)	0.22430							
									0.00047							
0.00989																
Figure 4f-h			Spearman's rank correlation coefficient	PA effect and learning attainment	0.74518	0.04664	0.00156	0.95715	0.1051	-0.5206	-0.8056	0.0286				
Figure 4j		Non-PA group = 13 mice PA group = 13 mice	One-way repeated measures ANOVA (post-hoc Bonferroni test)	(Pre-training) vs (0 day) (0 day) vs (+2 day) (0 day) vs (After PA) (+2 day) vs (After PA) (Pre-training) vs (After PA)	Non-PA (RotaRod)	Non-PA (Beam)	PA (RotaRod)	PA (Beam)								
					0.00000	0.00000	0.00000	0.00000								
					1.00000	NA	1.00000	NA								
					1.00000	1.00000	1.00000	0.04755								
					0.66982	NA	1.00000	NA								
					0.00000	0.00000	0.00000	0.00000								
Figure 4k			Spearman's rank correlation coefficient	PA effect on each task	0.06148			-0.53168								
Figure 5d, g, j, l, m		Dual task group = 1713 spines/5 mice Re-training group = 765 spines/5 mice Home cage group = 861 spines/5 mice	One-way factorial ANOVA (post-hoc Scheffe's test)	(Dual) vs (Re-training) (Re-training) vs (Homecage) (Dual) vs (Homecage)	Both Day2/4	Day 4 specific										
					0.09206	0.21206										
					0.00708	0.57905										
					0.35558	0.03786										
Figure 5e, h, k	In vivo M1 cortex + In utero EP (E14.5)	Dual task group: Day2 specific spines = 48, Both Day2/4 spines = 13, Day4 specific spines = 26 Re-training group: Day2 specific spines = 10, Both Day2/4 spines = 15, Day4 specific spines = 9 Home cage group: Day2 specific spines = 10, Both Day2/4 spines = 6, Day4 specific spines = 9	One-way factorial ANOVA (post-hoc Scheffe's test)	(Day2) vs (Both Day2/4) (Day2) vs (Day4) (Both Day2/4) vs (Day4)	Dual task	Re-training	Homecag									
					0.00331	0.04339	0.00147									
					0.00077	0.04579	0.01531									
					0.92817	0.95804	0.65233									
Figure 5n		Dual task group = 5 mice Re-training group = 5 mice Home cage group = 5 mice	Chi-squared test (Post-hoc Bonferonni correction)	(Dual) vs (Re-training) (Re-training) vs (Homecage) (Dual) vs (Homecage)	0.00000 1.00000 0.00000											

# Panorama of ancient metazoan macromolecular complexes

Cuihong Wan<sup>1,2\*</sup>, Blake Borgeson<sup>2\*</sup>, Sadhna Phanse<sup>1</sup>, Fan Tu<sup>2</sup>, Kevin Drew<sup>2</sup>, Greg Clark<sup>3</sup>, Xuejian Xiong<sup>4,5</sup>, Olga Kagan<sup>1</sup>, Julian Kwan<sup>1,4</sup>, Alexandr Bezginov<sup>3</sup>, Kyle Chessman<sup>4,5</sup>, Swati Pal<sup>5</sup>, Graham Cromar<sup>4,5</sup>, Ophelia Papoulas<sup>2</sup>, Zuyao Ni<sup>1</sup>, Daniel R. Boutz<sup>2</sup>, Snezana Stoilova<sup>1</sup>, Pierre C. Havugimana<sup>1</sup>, Xinghua Guo<sup>1</sup>, Ramy H. Malty<sup>6</sup>, Mihail Sarov<sup>7</sup>, Jack Greenblatt<sup>1,4</sup>, Mohan Babu<sup>6</sup>, W. Brent Derry<sup>4,5</sup>, Elisabeth R. Tillier<sup>3</sup>, John B. Wallingford<sup>2,8</sup>, John Parkinson<sup>4,5</sup>, Edward M. Marcotte<sup>2,8</sup> & Andrew Emili<sup>1,4</sup>

Macromolecular complexes are essential to conserved biological processes, but their prevalence across animals is unclear. By combining extensive biochemical fractionation with quantitative mass spectrometry, here we directly examined the composition of soluble multiprotein complexes among diverse metazoan models. Using an integrative approach, we generated a draft conservation map consisting of more than one million putative high-confidence co-complex interactions for species with fully sequenced genomes that encompasses functional modules present broadly across all extant animals. Clustering reveals a spectrum of conservation, ranging from ancient eukaryotic assemblies that have probably served cellular housekeeping roles for at least one billion years, ancestral complexes that have accrued contemporary components, and rarer metazoan innovations linked to multicellularity. We validated these projections by independent co-fractionation experiments in evolutionarily distant species, affinity purification and functional analyses. The comprehensiveness, centrality and modularity of these reconstructed interactomes reflect their fundamental mechanistic importance and adaptive value to animal cell systems.

## Introduction

Elucidating the components, conservation and functions of multi-protein complexes is essential to understand cellular processes<sup>1,2</sup>, but mapping physical association networks on a proteome-wide scale is challenging. The development of high-throughput methods for systematically determining protein–protein interactions (PPIs) has led to global molecular interaction maps for model organisms including *E. coli*, yeast, worm, fly and human<sup>3–10</sup>. In turn, comparative analyses have shown that PPI networks tend to be conserved<sup>11,12</sup>, evolve more slowly than regulatory networks<sup>13</sup>, and closely mirror function retention across orthologous groups<sup>11,14,15</sup>. Yet fundamental questions arise<sup>16,17</sup>. Here we define: (i) the extent to which physical interactions are preserved between phyla; (ii) the identity of protein complexes that are evolutionarily stable across animals; and (iii) the unique attributes of macromolecule composition, phylogenetic distribution and phenotypic significance.

## Generating a high-quality conserved interaction dataset

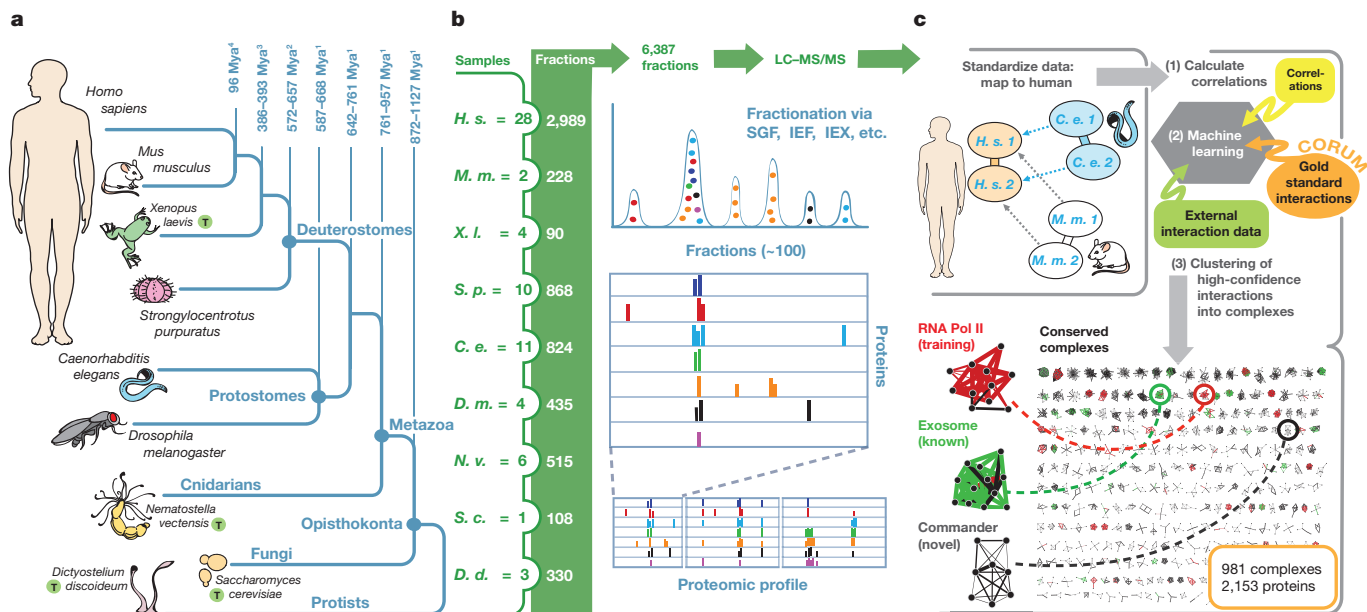
As previous cross-species interactome comparisons, based on experimental data from different sources and methods, show limited overlap<sup>12,18</sup>, we sought to produce a more comprehensive and accurate map of protein complexes common to metazoa by applying a standardized approach to multiple species. We employed biochemical fractionation of native macromolecular assemblies followed by tandem mass spectrometry to elucidate protein complex membership (Fig. 1; see Supplementary Methods). Previous application of this co-fractionation strategy to human cell lines preferentially identified vertebrate-specific protein complexes<sup>6</sup>, so we selected eight additional species for study on the basis of their relevance as model

organisms, spanning roughly a billion years of evolutionary divergence (Fig. 1a). The resulting co-fractionation data (Fig. 1b) acquired for *Caenorhabditis elegans* (worm), *Drosophila melanogaster* (fly), *Mus musculus* (mouse), *Strongylocentrotus purpuratus* (sea urchin), and human were used to discover conserved interactions (Fig. 1c), while the data obtained for *Xenopus laevis* (frog), *Nematostella vectensis* (sea anemone), *Dictyostelium discoideum* (amoeba) and *Saccharomyces cerevisiae* (yeast) were used for independent validation. Details on the cell types, developmental stages and fractionation procedures used are provided in Supplementary Table 1.

We identified and quantified (see Supplementary Methods) 13,386 protein orthologues across 6,387 fractions obtained from 69 different experiments (Fig. 2a), an order of magnitude expansion in data coverage relative to our original (*H. sapiens* only) study<sup>6</sup>. Individual pair-wise protein associations were scored based on the fractionation profile similarity measured in each species. Next, we used an integrative computational scoring procedure (Fig. 1c; see Supplementary Methods) to derive conserved interactions for human proteins and their orthologues in worm, fly, mouse and sea urchin, defined as high pair-wise protein co-fractionation in at least two of the five input species. The support vector machine learning classifier used was trained (using fivefold cross-validation) on correlation scores obtained for conserved reference annotated protein complexes (see Supplementary Methods), and combined all of the input species co-fractionation data together with previously published human<sup>6,19</sup> and fly interactions<sup>5</sup> and additional supporting functional association evidence<sup>20</sup> (HumanNet). Measurements of overall performance showed high precision with reasonable recall by the co-fractionation data alone (Fig. 2b), with external data sets serving only to increase

<sup>1</sup>Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, Toronto, Ontario M5S 3E1, Canada. <sup>2</sup>Center for Systems and Synthetic Biology, Institute for Cellular and Molecular Biology, University of Texas at Austin, Austin, Texas 78712, USA. <sup>3</sup>Department of Medical Biophysics, Toronto, Ontario M5G 1L7, Canada. <sup>4</sup>Department of Molecular Genetics, University of Toronto, Toronto, Ontario M5S 1A8, Canada. <sup>5</sup>Hospital for Sick Children, Toronto, Ontario M5G 1X8, Canada. <sup>6</sup>Department of Biochemistry, University of Regina, Regina, Saskatchewan S4S 0A2, Canada. <sup>7</sup>Max Planck Institute of Molecular Cell Biology and Genetics, 01307 Dresden, Germany. <sup>8</sup>Department of Molecular Biosciences, University of Texas at Austin, Austin, Texas 78712, USA.

\*These authors contributed equally to this work.



**Figure 1 | Workflow.** **a**, Phylogenetic relationships of organisms analysed in this study. We fractionated soluble protein complexes from worm (*C. elegans*) larvae, fly (*D. melanogaster*) S2 cells, mouse (*M. musculus*) embryonic stem cells, sea urchin (*S. purpuratus*) eggs and human (HEK293/HeLa) cell lines. Holdout species ("T", for test) likewise analysed were frog (*X. laevis*), an amphibian; sea anemone (*N. vectensis*), a cnidarian with primitive eumetazoan tissue organization; slime mould (*D. discoideum*), an amoeba; and yeast (*S. cerevisiae*), a unicellular eukaryote. **b**, Protein fractions were digested and

analysed by high-performance liquid chromatography tandem mass spectrometry (LC-MS/MS), measuring peptide spectral counts and precursor ion intensities. **c**, Integrative computational analysis. After orthologue mapping to human, correlation scores of co-eluting protein pairs detected in each 'input' species were subjected to machine learning together with additional external association evidence, using the CORUM complex database as a reference standard for training. High-confidence interactions were clustered to define co-complex membership.

precision and recall as we required all derived interactions to have extensive biochemical support (see Supplementary Methods). Co-fractionation data of each input species affected overall performance, in each case increasing precision and recall (Extended Data Fig. 1a). The final filtered interaction network consists of 16,655 high-confidence co-complex interactions in human (Supplementary Table 2). All of the interactions were supported by direct biochemical evidence in at least two input species, with half (8,121) detected in three or more (Extended Data Fig. 1b), enabling cross-species modelling and functional inference.

### Benchmarking protein complexes

Multiple lines of evidence support the quality of the network: reference complexes withheld during training were reconstructed with higher precision and recall (Fig. 2b; see Extended Data Fig. 1c) relative to our human-only map<sup>6</sup>. The interacting proteins were also sixfold enriched (hypergeometric  $P < 1 \times 10^{-24}$ ) for shared subcellular localization annotations in the Human Protein Atlas Database<sup>21</sup>, 21-fold enriched ( $P < 1 \times 10^{-56}$ ) for shared disease associations in OMIM<sup>22</sup>, and showed highly correlated human tissue proteome abundance profiles<sup>23</sup> (Extended Data Fig. 2a).

To independently verify the reliability of these projections, we examined the co-fractionation profiles of putatively interacting orthologues (interologues) in the four holdout species, as obtained by protein quantification across 1,127 biochemical fractions (see Supplementary Methods). Whereas sequence divergence changed absolute chromatographic retention times (Extended Data Fig. 2b), most of the predicted interactors showed highly correlated co-fractionation profiles among the holdout test species to a degree comparable to those of the input species used for learning (Fig. 2c). The biochemical data obtained for frog and sea anemone showed slightly better agreement than that for *Dictyostelium* and yeast that was proportional to evolutionary distance<sup>24</sup>.

Besides indicating stably associated proteins, our multispecies biochemical profiles faithfully recapitulated the architecture of

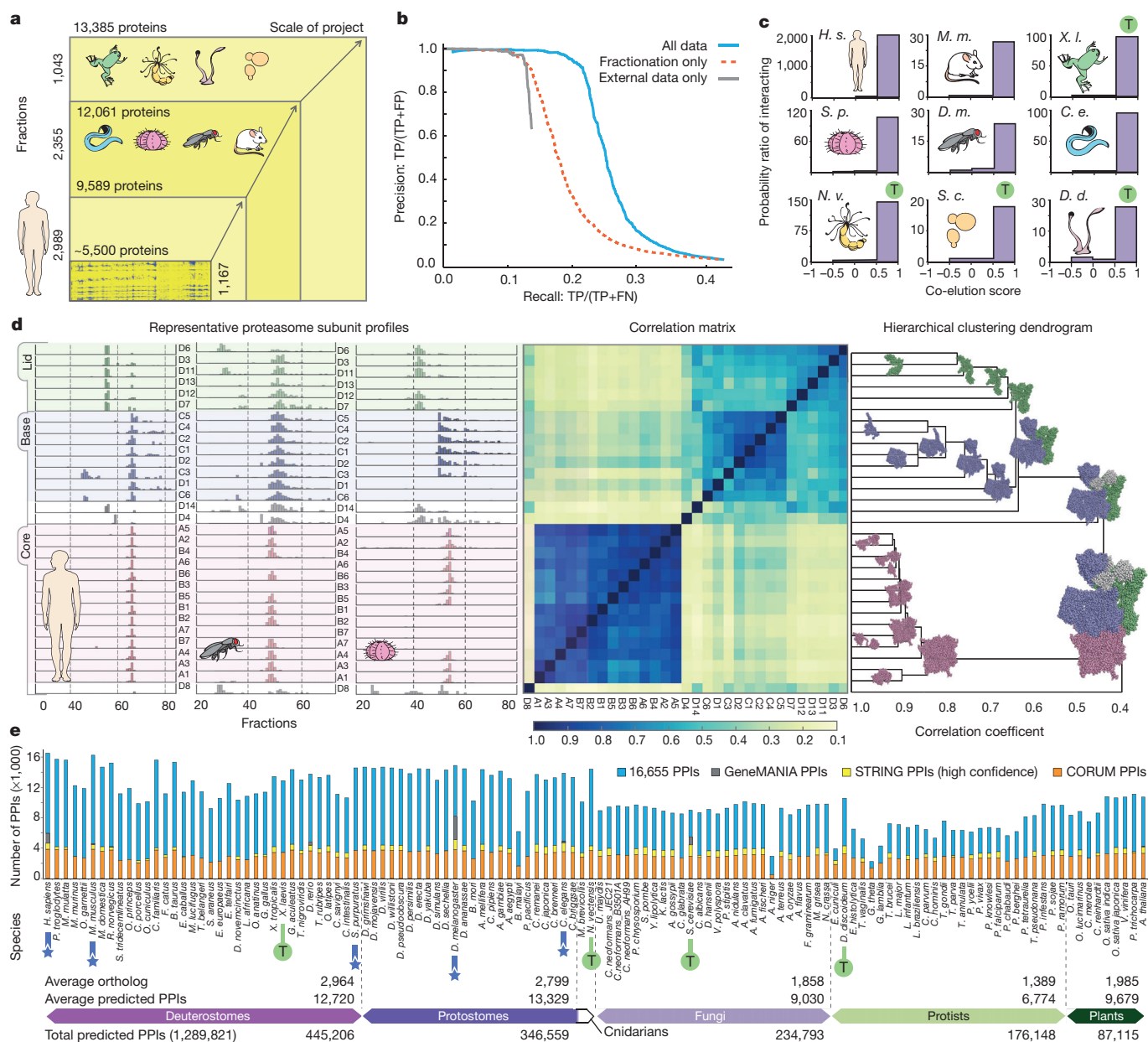
multiprotein complexes of known three-dimensional structure, with a general trend for most correlated protein pairs to be spatially closer (Extended Data Fig. 2c). For example, hierarchical clustering of 30S proteasome subunits according to chromatographic elution profiles of all five input species correctly separated the 20S and 19S particles and the regulatory lid from the base sub-complex (Fig. 2d), reflecting known hierarchies of complex formation and disassembly.

### Landscape of interaction conservation across species

Because most of the interacting components were phylogenetically conserved across vast evolutionary timescales, we were able to predict over one million high-confidence co-complex interactions among orthologous protein pairs for 122 extant eukaryotes with sequenced genomes (Supplementary Table 3). The number of interactions ranged from ~8,000 to ~15,000 per species depending on phyla (Fig. 2e), with more projected among Deuterostomes, Protostomes and Cnidaria, which show high component retention, and fewer in Fungi, Plants and, especially, Protists, where the relative paucity of co-complex conservation probably reflects inherent clade diversity, especially in parasite genomes (for example, gene loss among Apicomplexa). While largely congruent with previous smaller-scale studies of PPI conservation<sup>25</sup>, the majority of conserved co-complex interactions are novel (less than one-third curated in CORUM, STRING and GeneMANIA databases; Fig. 2e). This markedly increases the number of metazoan protein interactions reported to date (Supplementary Table 3), covering roughly 10%–25% of the estimated conserved animal cell interactome<sup>26,27</sup>, opening up many new avenues of inquiry.

To systematically define evolutionarily conserved functional modules, we partitioned the interaction network using a two-stage clustering procedure (Fig. 1c; see Supplementary Methods) that allowed proteins to participate in multiple complexes (that is, moonlighting) as merited (Extended Data Fig. 3a). The 981 putative multiprotein groupings (Fig. 3a; see Supplementary Table 4) include both





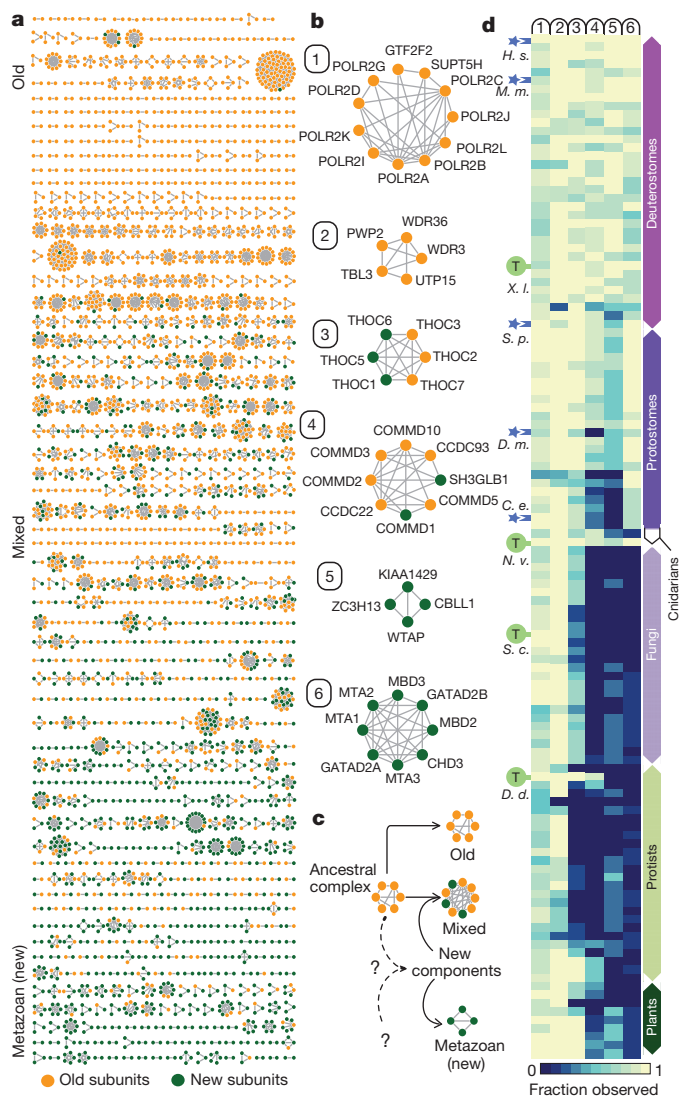
**Figure 2 | Derivation and projection of protein co-complex associations across taxa.** **a**, Expanded coverage via experimental scale-up relative to our previous human study<sup>6</sup>. Chart shows number of proteins detected, most (63%) in two or more species. **b**, Performance benchmarks, measuring precision and recall of our method and data in identifying known co-complex interactions (annotated human complexes from CORUM<sup>39</sup>). Complexes were split into training and withheld test sets; fivefold cross-validation against 4,528 interactions derived from the withheld test set shows strong performance gains, beyond baselines achieved using only co-fractionation or external evidence alone. TP, true positive; FP, false positive; FN, false negative. **c**, Plots showing high enrichment (probability ratio of interacting) of predicted interacting orthologous protein pairs (relative to non-interacting pairs) among highly

many well-known and novel complexes linked to diverse biological processes (Extended Data Fig. 3b). The complexes have estimated component ages spanning from ~500 million (metazoan-specific, or ‘new’) to over one billion years (ancient, or ‘old’) of evolutionary divergence. Details of species, orthologues, taxonomic groups, protein ages and evolutionary distances are provided in Supplementary Tables 3 and 5 and Supplementary Methods.

Although proteins arising in metazoa (by gene duplication or other means) account for about three quarters of all human gene products,

correlated fractionation profiles, in both the holdout validation (test, T) and input species (colours reflect clade memberships). **d**, Left, representative co-fractionation data (normalized spectral counts shown for portions of 3 of 42 experimental profiles) from human, fly and sea urchin showing characteristic profiles of proteasome core, base and lid sub-complexes. Hierarchical clustering (right) of pan-species pairwise Pearson correlation scores (centre) is consistent with accepted structural models (Protein Data Bank ID: 4CR2; core, red; base, blue; lid, green; out-clusters, white). **e**, Projection of conserved co-complex interactions across 122 eukaryotic species, indicating overlap with leading public PPI reference databases<sup>39–41</sup>. STRING bars indicate excess over CORUM; GeneMANIA bars indicate excess over both; component and interaction occurrences across clades indicated at bottom.

they form only about a third (39%; 147) of the clusters (Fig. 3a). These ‘new’ complexes tend to be smaller ( $\leq 3$  components; Fig. 3b) and specific (components not present in ‘mixed’ complexes). This indicates that although protein number and diversity greatly increased with the rise of animals<sup>25</sup>, most stable protein complexes were inherited from the unicellular ancestor and subsequently modified slightly over time (Fig. 3c and Supplementary Table 5). Indeed, the dominant phylogenetic profile of complexes across Eukarya (Fig. 3d) is composed either entirely (344 old complexes) or predominantly (490



**Figure 3 | Prevalence of conservation of protein complexes across Metazoa and beyond.** **a**, Conserved multiprotein complexes, identified by clustering, arranged according to average estimated component age (see Supplementary Methods and ref. 25). Proteins (nodes) classified as metazoan (green) or ancient (orange); assemblies showing divergent phylogenetic trajectories termed 'mixed'. **b**, Example complexes with different proportions of old and new subunits. **c**, Presumed origins of metazoan (new), mixed and old complexes; '?' indicates variable origins of new genes. **d**, Heat map showing prevalence of selected complexes across phyla. Colour reflects fraction of components with detectable orthologues (absence, dark blue). Sea anemone (*N. vectensis*) is the most distant metazoan (cnidarian) analysed biochemically.

mixed complexes) of ancient subunits ubiquitous among eukaryotes (Extended Data Fig. 4a; see Supplementary Table 5 for details), the latter presumably reflecting preferential accretion of additional components to pre-existing macromolecules (Fig. 3c)<sup>28</sup>.

These primordial complexes are present throughout the Opisthokonta supergroup (animals and fungi), estimated to be more than one billion years old<sup>29</sup>, and plants (and presumably lost/significantly diverged among parasitic protists). Reflecting this central importance, these complexes tend strongly to be ubiquitously expressed throughout all cell types and tissues (Extended Data Fig. 5a), are abundant (Extended Data Fig. 5b), and are enriched for associations to human disease and perturbation phenotypes in *C. elegans* (Supplementary Table 6). In comparison with other proteins in the 16,655 interactions, the older, conserved proteins present in these stable complexes have lower average domain complexity

( $P < 0.02$ ; see Supplementary Methods), suggesting multi-domain architectures underlie more transient or tissue-specific interactions. Whereas mixed and old complexes are enriched for functional associations with core cellular processes, such as metabolism (Extended Data Fig. 4c), the strictly metazoan complexes were far more likely to be linked to cell adhesion, organization and differentiation, consistent with roles in multicellularity. Reflecting these different evolutionary trajectories, new clusters are substantially more enriched for cancer-related proteins (42%; 62/147; hypergeometric  $P \leq 1 \times 10^{-5}$ ) compared to strictly old (15%; 53/344;  $P \leq 1 \times 10^{-3}$ ) clusters (Z-test  $< 0.0001$ ) (Supplementary Table 7), have generally lower annotation rates (Extended Data Fig. 4b), and show different preponderances of protein domains (Extended Data Fig. 4c and Supplementary Table 6).

### Independent biological assessment

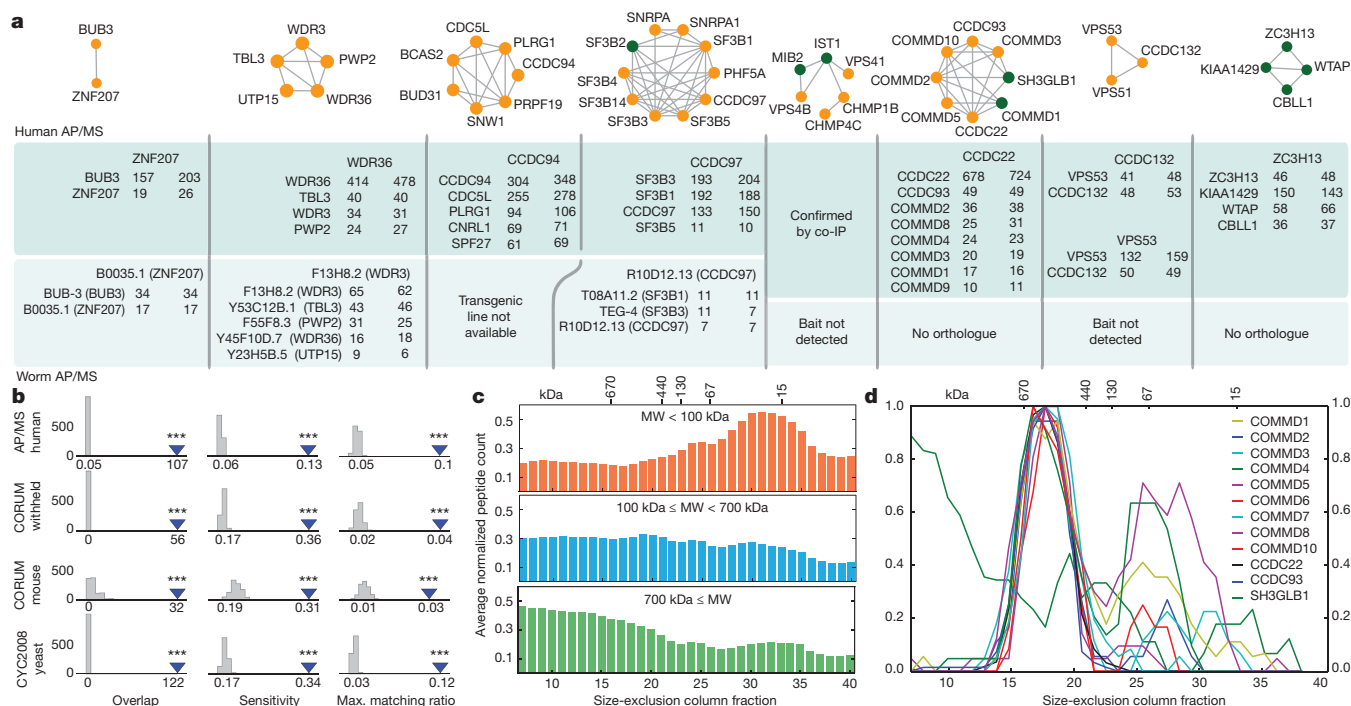
We used multiple approaches to assess the accuracy (Fig. 4) and functional significance (Fig. 5) of the predicted complexes. First, we performed affinity purification mass spectrometry (AP/MS) experiments on select novel complexes from the new, old and mixed age clusters, validating most associations in both worm and human (Fig. 4a and Extended Data Fig. 6a). We next performed a global validation by comparing our derived complexes to a newly reported large-scale AP/MS study of 23,756 putative human protein interactions detected in cell culture (E. L. Huttlin *et al.*, BioGRID preprint 166968), and observed a partial, but highly statistically significant, overlap to a degree comparable to literature-derived complexes (Fig. 4b, Extended Data Fig. 6b).

We also observed broad agreement between the derived complexes' inferred molecular weights (assuming 1:1 stoichiometries) and migration by size-exclusion chromatography (Fig. 4c and Extended Data Fig. 7a) and density gradient centrifugation (Extended Data Fig. 7b). A prime example is the coherent profiles of a large (~500 kDa) mixed complex with several un-annotated components (Fig. 4d and Extended Data Fig. 8), dubbed 'Commander', because most subunits share COMM (copper metabolism MURR1) domains<sup>30</sup> implicated in copper toxicosis<sup>31</sup>, among other roles<sup>30,32</sup>. Commander contains coiled-coil domain proteins CCDC22 and CCDC93 (Figs 4a, d) in addition to ten COMM domain proteins, broadly supported by co-fractionation in human, fly and sea urchin (Extended Data Fig. 9a–c and supporting website, [http://metazoa.med.utoronto.ca/php/view\\_elution\\_image.php?id=71&cond=ms2](http://metazoa.med.utoronto.ca/php/view_elution_image.php?id=71&cond=ms2)).

We found an unexpected role in embryonic development for Commander, whose subunits are strongly co-expressed in developing frog (Extended Data Fig. 9d, e). *COMM2/3*-knockdown (morpholino) tadpoles showed impaired head and eye development (Fig. 5a and Extended Data Fig. 9f, h), and defective neural patterning and expression changes in brain markers *PAX6*, *EN2* and *KROX20/EGFR1* (Fig. 5b and Extended Data Fig. 9g, h). Given the recently discovered link<sup>33,34</sup> between *CCDC22* and human syndromes of intellectual disability, malformed cerebellum and craniofacial abnormalities, the deep conservation of the Commander complex suggests *COMM2/3* as strong candidates in the aetiology of these heterogeneous disorders.

Among metazoan-specific protein complexes, we confirmed physical and functional associations of spindle checkpoint protein BUB3 with ZNF207, a zinc-finger protein conspicuously lacking orthologues in cnidarians and fungi. ZNF207 binds Bub3 via a Gle2-binding-sequence (GLEBS) motif<sup>35</sup> restricted to deuterostomes and protostomes (Extended Data Fig. 10a). As in human, knockdown of the ZNF207 orthologue in *C. elegans* (*B0035.1*) enhanced lethality owing to impaired Bub3-mediated checkpoint arrest (Fig. 5c).

Among mixed complexes, we confirmed metazoan-specific coiled-coil domain protein CCDC97 as a sub-stoichiometric component of human and worm SF3B spliceosomal complex involved in branch-site recognition (Fig. 4a). Consistent with a possible role in



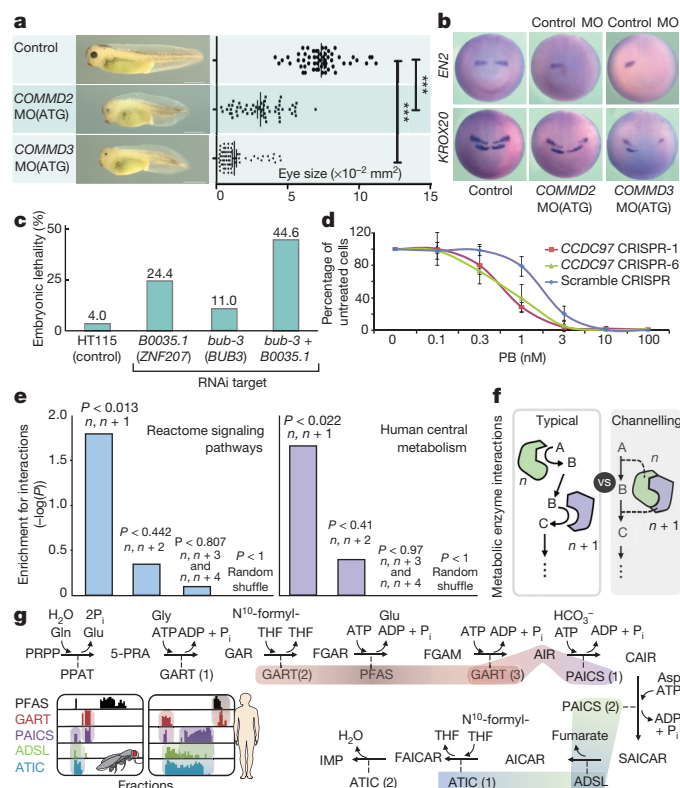
**Figure 4 | Physical validation of complexes.** **a**, Verification of complexes from tagged human cell lines and transgenic worms (see Supplementary Methods; complexes drawn as in Fig. 3). Inset reports spectral counts obtained in replicate AP/MS analyses of indicated bait protein (header). MIB2–VPS4 complex confirmed by co-immunoprecipitation (co-IP; Extended Data Fig. 6a). **b**, Conserved complexes significantly overlap large-scale AP/MS data reported for human cell lines (E. L. Huttlin *et al.*, BioGRID preprint 166968) to a

comparable extent as literature reference sets<sup>39,42</sup>, using three measures of complex-level agreement (see Supplementary Methods, Extended Data Fig. 6b); \*\*\* $P < 0.001$ , determined by shuffling (grey distributions). **c**, Agreement of inferred molecular weights (MW) of human protein complexes with size-exclusion chromatography profiles (data in **c**, **d**, from ref. 43). **d**, Co-elution of human Commander complex subunits by size-exclusion chromatography consistent with an approximately 500-kDa particle.

pre-mRNA splicing, CRISPR-based *CCDC97*-knockout human cells were slower growing than were control lines (Extended Data Fig. 10b, c) and hypersensitive to pladienolide B (Fig. 5d), a macrolide inhibitor of SF3b<sup>36</sup>.

## Network perspective into conserved biological systems

Knowledge of conserved macromolecular associations provides a road map for additional functional inferences. For instance, fractionation profiles can be compared for any pair of proteins in our data set to search for evidence of interactions. We found significant enrichment for interactions among pairs of human proteins acting sequentially in annotated pathways<sup>37</sup> (Fig. 5e), especially G-protein and MAP-kinase cascades (Supplementary Table 8). Enzymes acting consecutively in core metabolic reactions (Fig. 5f) also showed a higher tendency to interact (Supplementary Table 8), the significance of which decayed with more intervening steps (Fig. 5e). For example, strong consecutive



**Figure 5 | Functional validation of complexes.** **a**, Morpholino (MO(ATG), targeting start codon to block translation) knockdown of *COMMD2* ( $n = 55$  animals, 2 clutches, 1 eye each) or *COMMD3* ( $n = 64$ ) in *X. laevis* embryos causes defective head and eye development (control  $n = 57$ ; Extended Data Fig. 9f, h). \*\*\* $P < 0.0001$ , 2-sided Mann–Whitney test. **b**, *COMMD2/3* knockdown animals (five embryos per treatment examined) show altered neural patterning, including posterior shift or loss of expression of mid-brain marker *EN2* and *KROX20* (*EGRI*), the latter in rhombomeres R3/R5 (compare to Extended Data Fig. 9g, h). **c**, Enhanced embryonic lethality (epistasis) following RNAi knockdown in *C. elegans* of *B0035.1* (*ZNF207*) and *bub-3* together (eggs laid: HT115, 1,308; *B0035.1*, 1,096; *bub-3*, 445; *bub-3* + *B0035.1*, 341). **d**, Enhanced sensitivity (mean  $\pm$  s.d. across four cell culture experiments) of two independent *CCDC97*-knockout lines to the SF3b inhibitor pladienolide B (PB) relative to control HEK293 cells. **e**, Enrichment (permutation test  $P$  value) for interactions among sequential pathway components and metabolic enzymes relative to shuffled controls ( $n$  refers to enzyme index, where  $n, n + 1$  denotes sequential enzymes,  $n, n + 2$  sequential-but-one, and so on, as described in Supplementary Information. **f**, Metabolic channelling as opposed to traditional (typical) two-step cascade model. **g**, Conserved interactions among consecutively acting enzymes involved in purine biosynthesis (two representative co-fractionation profiles of the 69 total generated are shown).



interactions were apparent within the widely conserved purine biosynthetic pathway, with enzymes (for example, PAICS, GART) eluting in two peaks (Fig. 5g), one coincident with the prior enzyme and the second with the downstream enzyme, suggestive of substrate channelling<sup>38</sup>.

Despite the diversity of multicellular organisms, our study reveals fundamental attributes of the macromolecular machinery of animal cells with near universal pertinence to metazoan biology, development and evolution. Our extremely large set of supporting biochemical fractionation data (via ProteomeXchange with identifiers PXD002319–PXD002328), PPIs (via BioGRID; <http://thebiogrid.org/185267/publication/>) and interaction network projections are fully accessible (<http://metazoa.med.utoronto.ca>) to facilitate in-depth exploration. Although we focused on global conservation properties, these data can be analysed at the individual animal species or complex levels to assess the variety and functional adaptations of particular protein assemblies across phyla.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 15 December 2014; accepted 30 June 2015.**

**Published online 7 September 2015.**

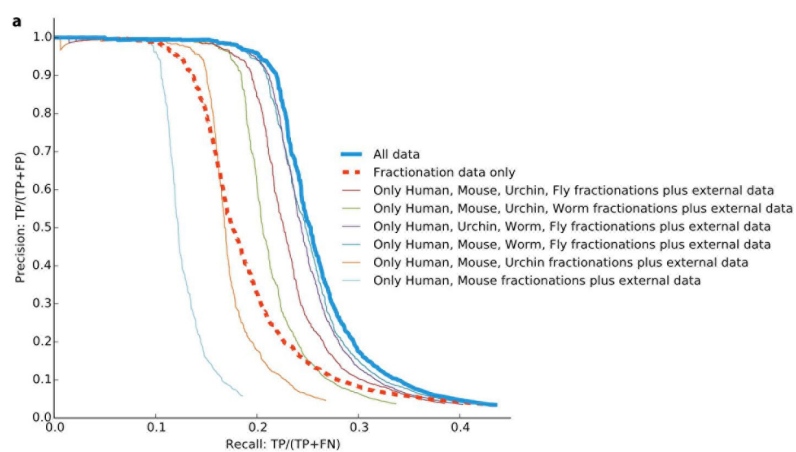
- Hartwell, L. H., Hopfield, J. J., Leibler, S. & Murray, A. W. From molecular to modular cell biology. *Nature* **402**, C47–C52 (1999).
- Alberts, B. The cell as a collection of protein machines: Preparing the next generation of molecular biologists. *Cell* **92**, 291–294 (1998).
- Butland, G. et al. Interaction network containing conserved and essential protein complexes in *Escherichia coli*. *Nature* **433**, 531–537 (2005).
- Krogan, N. J. et al. Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*. *Nature* **440**, 637–643 (2006).
- Gururharsha, K. G. et al. A protein complex network of *Drosophila melanogaster*. *Cell* **147**, 690–703 (2011).
- Havugimana, P. C. et al. A census of human soluble protein complexes. *Cell* **150**, 1068–1081 (2012).
- Stelzl, U. et al. A human protein-protein interaction network: a resource for annotating the proteome. *Cell* **122**, 957–968 (2005).
- Li, S. et al. A map of the interactome network of the metazoan *C. elegans*. *Science* **303**, 540–543 (2004).
- Hu, P. et al. Global functional atlas of *Escherichia coli* encompassing previously uncharacterized proteins. *PLoS Biol.* **7**, e1000096 (2009).
- Rolland, T. et al. A proteome-scale map of the human interactome network. *Cell* **159**, 1212–1226 (2014).
- Sharan, R. et al. Conserved patterns of protein interaction in multiple species. *Proc. Natl Acad. Sci. USA* **102**, 1974–1979 (2005).
- Gandhi, T. K. B. et al. Analysis of the human protein interactome and comparison with yeast, worm and fly interaction datasets. *Nature Genet.* **38**, 285–293 (2006).
- Tan, K., Shlomi, T., Feizi, H., Ideker, T. & Sharan, R. Transcriptional regulation of protein complexes within and across species. *Proc. Natl Acad. Sci. USA* **104**, 1283–1288 (2007).
- Singh, R., Xu, J. B. & Berger, B. Global alignment of multiple protein interaction networks with application to functional orthology detection. *Proc. Natl Acad. Sci. USA* **105**, 12763–12768 (2008).
- Yu, H. et al. Annotation transfer between genomes: protein–protein interologs and protein–DNA regulogs. *Genome Res.* **14**, 1107–1118 (2004).
- Ideker, T. & Krogan, N. J. Differential network biology. *Mol. Syst. Biol.* **8**, 565 (2012).
- Kiemer, L. & Cesareni, G. Comparative interactomics: comparing apples and pears? *Trends Biotechnol.* **25**, 448–454 (2007).
- von Mering, C. et al. Comparative assessment of large-scale data sets of protein–protein interactions. *Nature* **417**, 399–403 (2002).
- Malovannaya, A. et al. Analysis of the human endogenous coregulator complexome. *Cell* **145**, 787–799 (2011).
- Lee, I., Blom, U. M., Wang, P. I., Shim, J. E. & Marcotte, E. M. Prioritizing candidate disease genes by network-based boosting of genome-wide association data. *Genome Res.* **21**, 1109–1121 (2011).
- Uhlen, M. et al. Towards a knowledge-based Human Protein Atlas. *Nature Biotechnol.* **28**, 1248–1250 (2010).
- McKusick, V. A. *Mendelian Inheritance in Man: A Catalog of Human Genes and Genetic Disorders*. (Johns Hopkins Univ. Press, 1998).
- Kim, M. S. et al. A draft map of the human proteome. *Nature* **509**, 575–581 (2014).
- Rubin, G. M. et al. Comparative genomics of the eukaryotes. *Science* **287**, 2204–2215 (2000).
- Bezginov, A., Clark, G. W., Charlebois, R. L., Dar, V. U. N. & Tillier, E. R. M. Coevolution reveals a network of human proteins originating with multicellularity. *Mol. Biol. Evol.* **30**, 332–346 (2013).
- Stumpf, M. P. H. et al. Estimating the size of the human interactome. *Proc. Natl Acad. Sci. USA* **105**, 6959–6964 (2008).
- Hart, G. T., Ramani, A. K. & Marcotte, E. M. How complete are current yeast and human protein–interaction networks? *Genome Biol.* **7**, 120 (2006).
- Eisenberg, E. & Levanon, E. Y. Preferential attachment in the protein network evolution. *Phys. Rev. Lett.* **91**, 138701 (2003).
- Knoll, A. H. The early evolution of eukaryotes: a geological perspective. *Science* **256**, 622–627 (1992).
- Burstein, E. et al. COMMD proteins, a novel family of structural and functional homologs of MURR1. *J. Biol. Chem.* **280**, 22222–22232 (2005).
- van de Sluis, B., Rothuizen, J., Pearson, P. L., van Oost, B. A. & Wijmenga, C. Identification of a new copper metabolism gene by positional cloning in a purebred dog population. *Hum. Mol. Genet.* **11**, 165–173 (2002).
- McDonald, F. J. COMMD1 and ion transport proteins: what is the COMMD1? Focus on “COMMD1 interacts with the COOH terminus of NKCC1 in Calu-3 airway epithelial cells to modulate NKCC1 ubiquitination”. *Am. J. Physiol. Cell Physiol.* **305**, C129–C130 (2013).
- Kolanczyk, M. et al. Missense variant in CCDC22 causes X-linked recessive intellectual disability with features of Ritscher-Schinzel/3C syndrome. *Eur. J. Hum. Genet.* **109**, 1–6 (2014).
- Voineagu, I. et al. CCDC22: a novel candidate gene for syndromic X-linked intellectual disability. *Mol. Psychiatry* **17**, 4–7 (2012).
- Toledo, C. M. et al. BuGZ is required for Bub3 stability, Bub1 kinetochore function, and chromosome alignment. *Dev. Cell* **28**, 282–294 (2014).
- Kotake, Y. et al. Splicing factor SF3b as a target of the antitumor natural product pladienolide. *Nature Chem. Biol.* **3**, 570–575 (2007).
- Croft, D. et al. The Reactome pathway knowledgebase. *Nucleic Acids Res.* **42**, D472–D477 (2014).
- Ovádi, J. *Cell Architecture and Metabolite Channeling*. (RG Landes Company, 1995).
- Ruepp, A. et al. CORUM: the comprehensive resource of mammalian protein complexes–2009. *Nucleic Acids Res.* **38**, D497–D501 (2010).
- Warde-Farley, D. et al. The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Res.* **38**, W214–W220 (2010).
- Franceschini, A. et al. STRING v9.1: protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Res.* **41**, D808–D815 (2013).
- Pu, S., Wong, J., Turner, B., Cho, E. & Wodak, S. J. Up-to-date catalogues of yeast protein complexes. *Nucleic Acids Res.* **37**, 825–831 (2009).
- Kirkwood, K. J., Ahmad, Y., Larance, M. & Lamond, A. I. Characterization of native protein complexes and protein isoform variation using size-fractionation-based quantitative proteomics. *Mol. Cell. Proteomics* **12**, 3851–3873 (2013).

**Supplementary Information** is available in the online version of the paper.

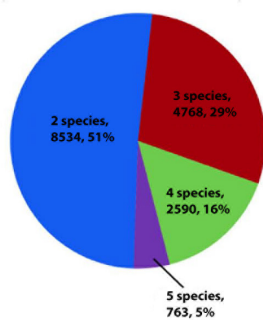
**Acknowledgements** We thank G. Bader, P. Kim, G. Moreno-Hagelsieb, S. Pu and S. Wodak for critical suggestions, illustrator A. Syrett for expert help drafting figures, T. Kwon (University of Texas) for *X. laevis* gene models, and K. Foltz (University of California, Santa Barbara), A. Brehm (Philipps-University Marburg), P. Paddison (Fred Hutchinson Cancer Research Center), J. Smith (Woods Hole Marine Biological Laboratory), P. Zandstra and J. Moffat (University of Toronto) for providing biological specimens and reagents. We thank members of the Emili and Marcotte laboratories for assistance and guidance, and SciNet (University of Toronto) and the Texas Advanced Computing Center (University of Texas) for high-performance computing resources. This work was supported by grants from the CIHR, NSERC, ORF and the CFI to A.E., from the CIHR and Heart and Stroke to J. P., from the NIH (F32GM112495) to K.D., and from the NIH, NSF, CPRIT, and Welch Foundation (F-1515) to E.M.M.

**Author Contributions** A.E. and E.M.M. designed and co-supervised the project. C.W. performed proteomic experiments, aided by P.C.H. B.B. coordinated data analysis, aided by S.Ph., K.D. and S.S., and guided by E.M.M. E.R.T., G.Cl., A.B., J.P., X.X., K.C., C.W. and S.Ph. analysed network and conservation data. C.W., F.T., O.K., J.K., S.Pa., O.P., Z.N., D.R.B., X.G., R.H.M., M.S., J.G., M.B., W.B.D. and J.B.W. contributed validation experiments. S.Ph. designed the web portal. C.W., B.B., E.M.M. and A.E. drafted the manuscript. All authors discussed results and contributed edits.

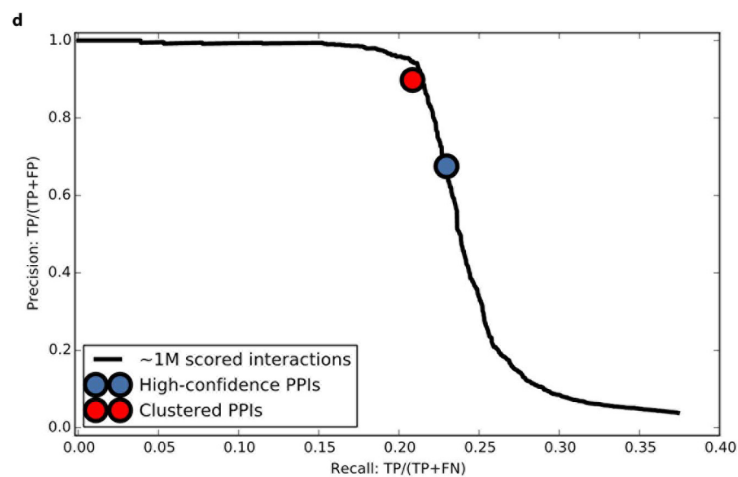
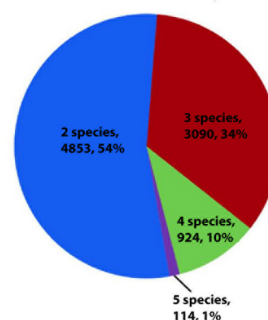
**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to E.M.M. ([marcotte@icmb.utexas.edu](mailto:marcotte@icmb.utexas.edu)) or A.E. ([andrew.emili@utoronto.ca](mailto:andrew.emili@utoronto.ca)).



**b** Proportion of PPI across species



**c** Novel PPI across species



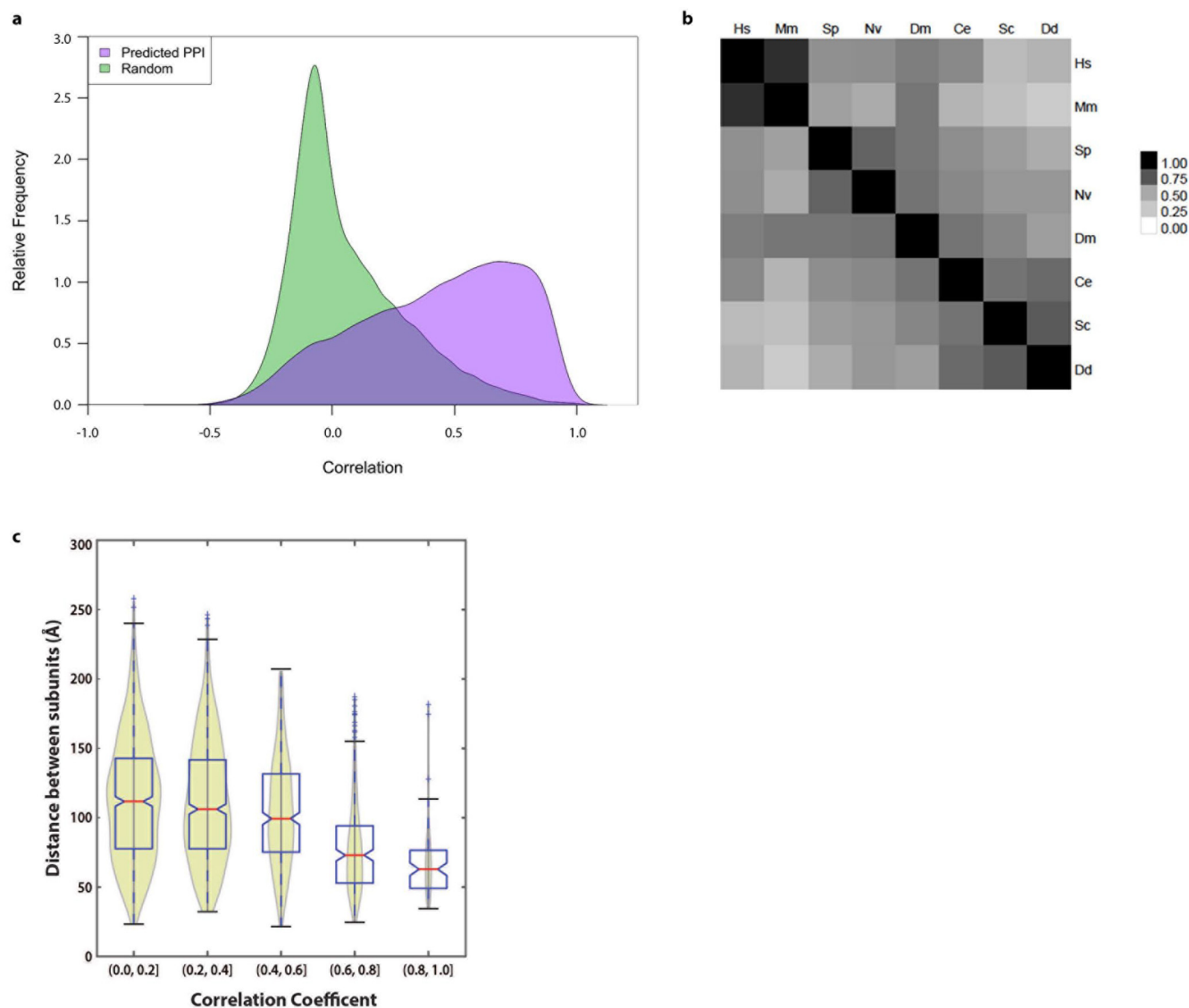
**Extended Data Figure 1 | Performance measures.** **a**, Performance benchmarks, measuring the precision and recall of our method and data in identifying known co-complex interactions from a withheld reference set of annotated human complexes (from CORUM<sup>39</sup>; as in Fig. 2b). Fivefold cross-validation against this withheld set shows strong performance gains, beyond a baseline achieved using only human and mouse co-fractionation data along with additional evidence from independent protein interaction screens<sup>5,19</sup> and a functional gene network<sup>20</sup> (far-left curve), made by integrating co-fractionation data from the additional non-human animal species (as indicated). ‘All data’ and ‘Fractionation data only’ curves include biochemical fractionation data from all five input species: human, mouse, urchin, fly and worm; the latter curve omits all external data. In all cases, at least two species were required to show supporting biochemical evidence. Recall refers to the fraction of 4,528 total positive interactions derived from the withheld human CORUM complexes. **b**, All 16,655 interactions were identified at least in two species, half (49%, 8,121) found in three or more species. **c**, Among these high-confidence co-complex interactions, 8,981 (54%) were not reported in iRefWeb<sup>44</sup> (v13.0), BioGRID<sup>45</sup> (v3.2.119) or CORUM reference

(Supplementary Table 2) for any of the five input species or in yeast; half (46%, 4,128) of these novel co-complex interactions display evidence of co-fractionation in three or more species. **d**, Final precision/recall performance on withheld interaction test set. A support vector machine classifier was trained using interactions derived from our training set of CORUM complexes, then ~1 million protein pairs found to co-elute in at least two of the five input species were scored by the classifier. Black curve shows precision and recall for ranked list of co-eluting pairs, with recall representing the fraction recovered of 4,528 total positive interactions derived from the withheld set of merged human CORUM complexes, and precision measured using co-eluting pairs where both members of the pair are contained in the set of proteins represented in the CORUM withheld set. The top 16,655 pairs, giving a cumulative precision of 67.5% and recall of 23.0% on this withheld test set, form the high-confidence set of co-complex protein–protein interactions (blue circle). The highest-scoring interactions were clustered using the two-stage approach described in the Supplementary Methods, yielding a final set of 7,669 interactions, which form the 981 identified complexes (red circle; precision = 90.0%, recall = 20.8%).

44. Turner, B. *et al.* iRefWeb: interactive analysis of consolidated protein interaction data and their supporting evidence. *Database* **2010**, baq023 (2010).

45. Stark, C. *et al.* BioGRID: a general repository for interaction datasets. *Nucleic Acids Res.* **34**, D535–D539 (2006).

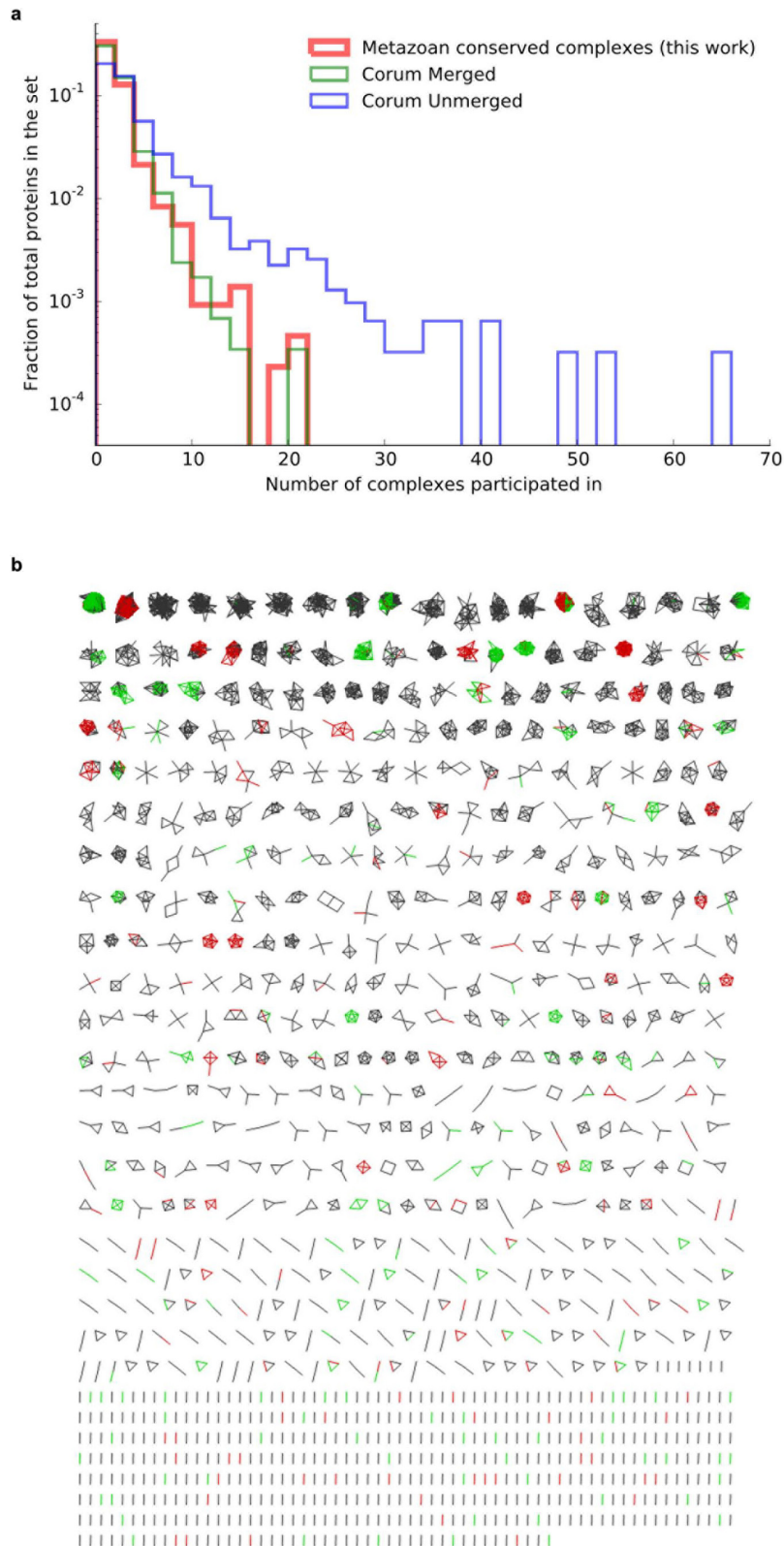




### Extended Data Figure 2 | Properties of protein elution profiles.

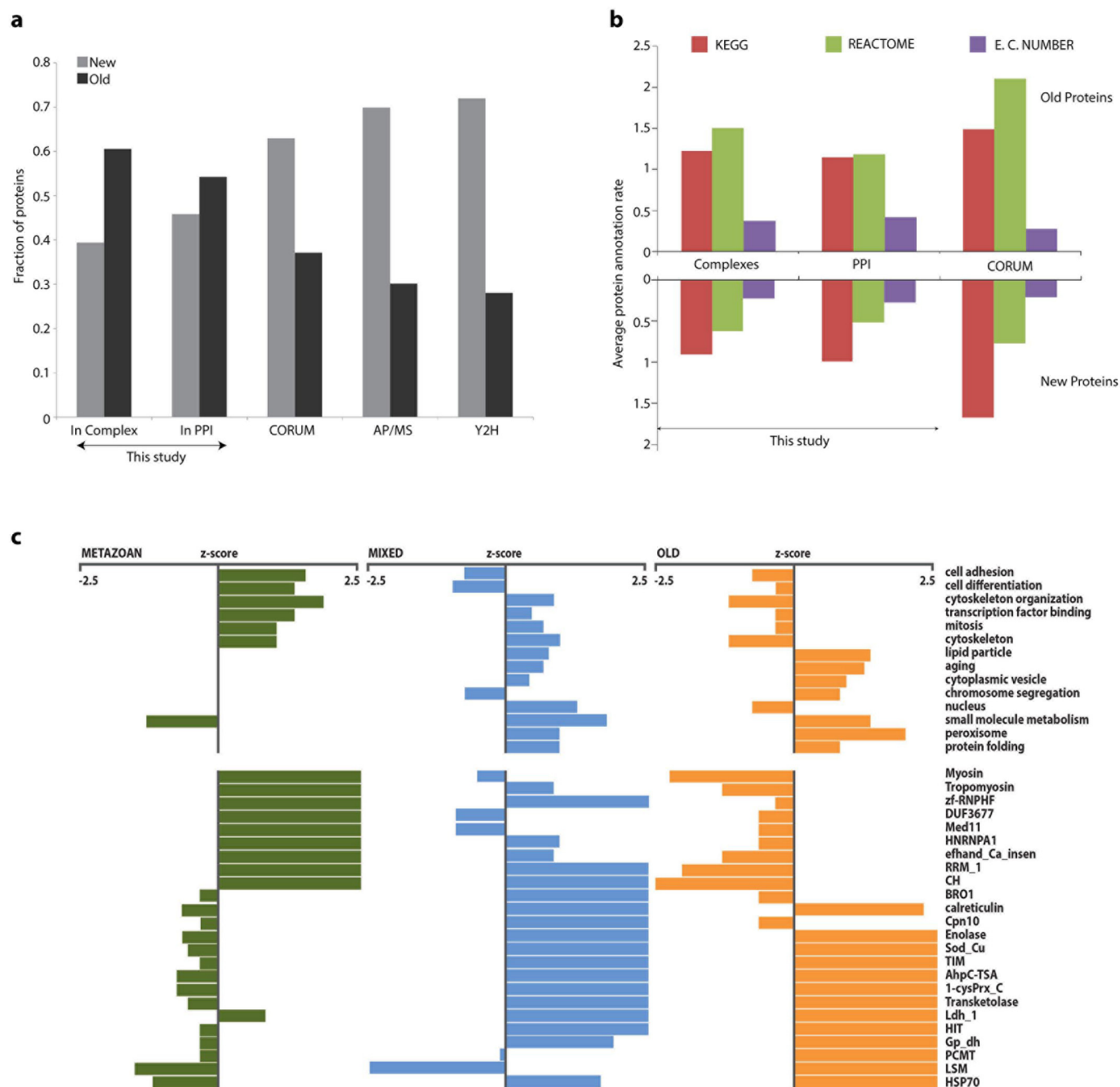
**a**, Distribution of global protein tissue expression pattern similarity, measured as the Pearson correlation coefficient of protein abundance across 30 human tissues<sup>23</sup>, showing markedly higher correlations for 16,468 protein–protein pairs of putative co-complex interaction partners compared to the same number of randomized pairs of proteins in the network which were not predicted to interact. **b**, Heat map illustrating the low to moderate cross-species Spearman's rank correlation coefficients in the elution profiles observed between orthologous proteins during mixed-bed ion exchange

chromatography under standardized conditions, highlighting the shift in absolute chromatographic retention times in different species. This variation indicates that the conservation of co-fractionation by putatively interacting proteins is not merely a trivial result stemming from fixed column-retention times. **c**, The degree of co-fractionation is measured as the correlation coefficient between elution profiles. Spatial proximity is calculated from the mean of residue pair distances between components of multisubunit complexes with known three-dimensional structures (see Supplementary Methods).



**Extended Data Figure 3 | Derivation of complexes.** **a**, The 2,153 proteins present in the 981 derived metazoan complexes participate in multiple assemblies ('moonlighting') to an extent comparable to the sharing of subunits reported for literature-derived complexes (CORUM). For comparison, we examined the 1,550 unique proteins from the full CORUM set of 1,216 human complexes passing our selection criteria for supporting evidence ('Unmerged') and the 1,461 unique proteins from the non-redundant set of 501 merged complexes used as the reference for splitting our training and testing sets, with some of the largest complexes removed to avoid bias in training ('Merged');

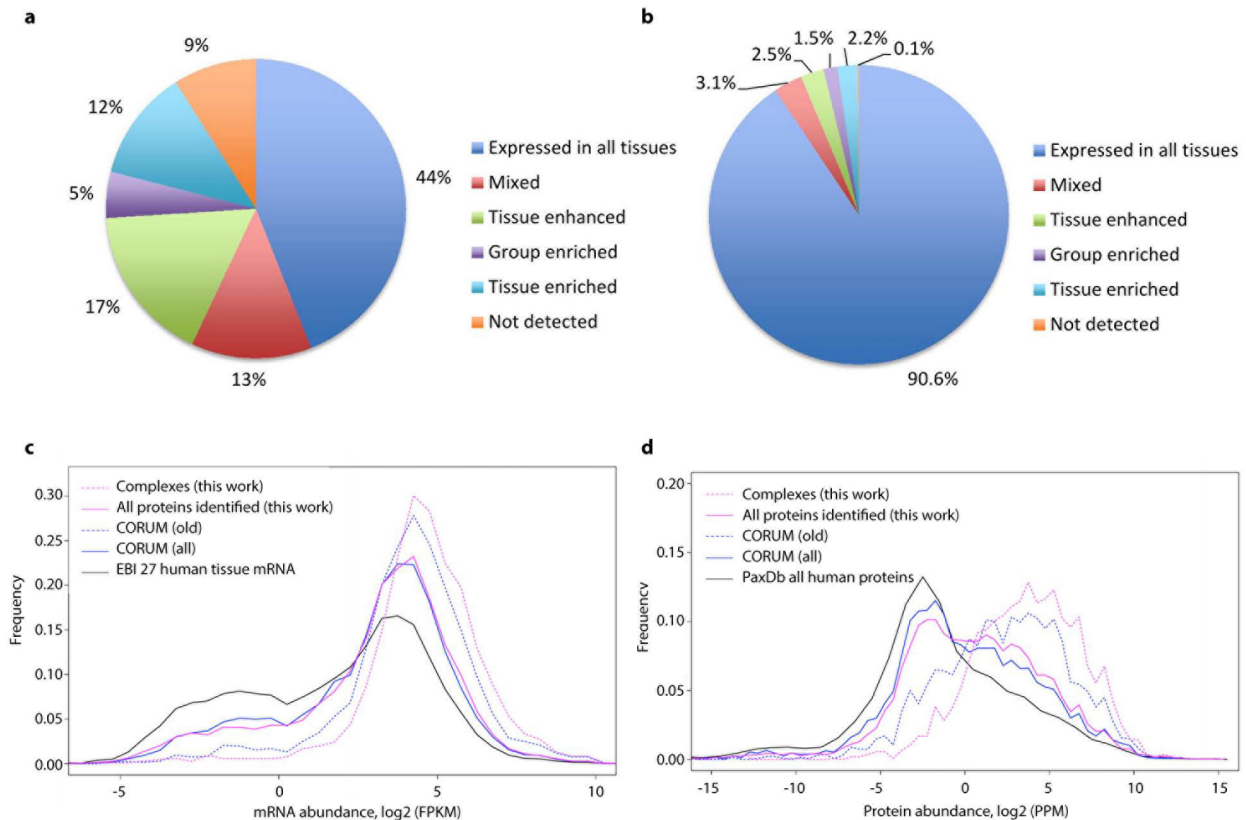
see 'Optimizing the two-stage clustering' in Supplementary Methods for details). **b**, Schematic of 981 identified complexes containing 2,153 unique proteins. In this graphical representation, 7,669 co-complex interactions are shown as lines, and proteins as nodes. Red and green interactions were previously annotated in CORUM. Red interactions were used in training the classifier and/or clustering procedure, while green interactions were held out for validation purposes. Grey interactions were not previously annotated in CORUM.



**Extended Data Figure 4 | Properties of new and old proteins and complexes.** **a**, The 2,153 protein components in the conserved animal complexes tend to be more ancient than the 2,301 proteins reported in the CORUM reference complexes or in two recent large-scale protein interaction assays, based on either the 7,062 proteins found by affinity purification/mass spectrometry (AP/MS; E. L. Huttlin *et al.*, BioGRID preprint 166968, <http://thebiogrid.org/166968/publication/>) or the 3,667 proteins analysed by yeast two-hybrid assays (Y2H)<sup>10</sup>. Ages are derived from OMA (Orthologous Matrix

database) as in ref. 25. **b**, Annotation rates (mean count of annotation terms per protein) of old and new proteins in the derived complexes and pairwise PPIs, compared with proteins in the CORUM reference complex set. Old proteins (defined by OMA) from the complexes generally exhibited higher annotation rates than new proteins. **c**, Differential enrichment of old, mixed and metazoan-specific protein complexes for functional annotations (select GO-slim biological process terms shown, top) and protein domains (Pfam, bottom).

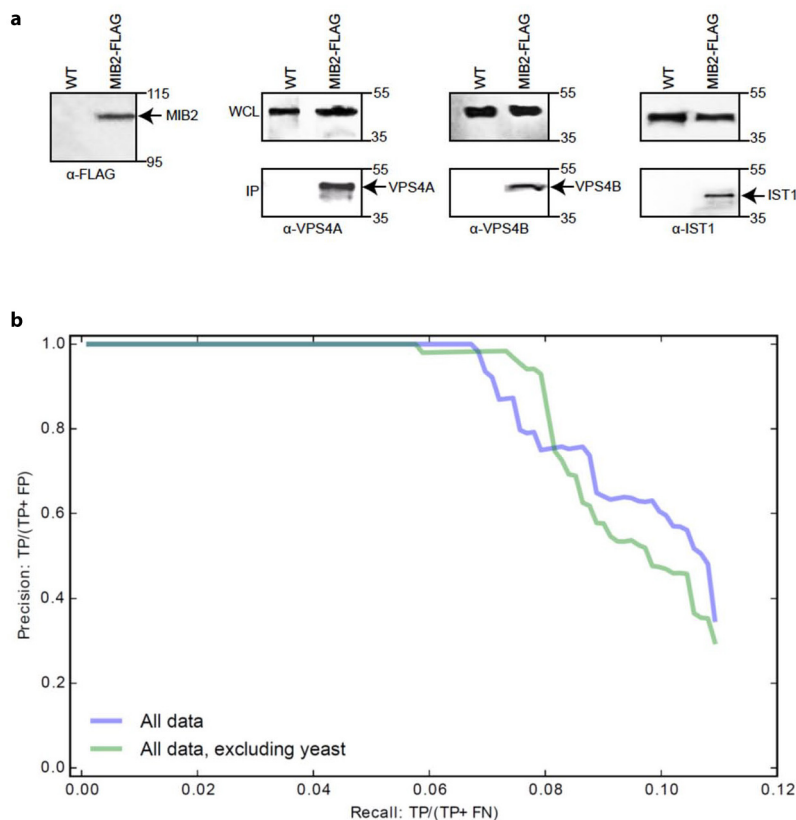




**Extended Data Figure 5 | Abundance and expression trends for proteins in complexes.** Proteins within the identified complexes tend to be ubiquitously expressed across human tissues. **a, b**, Pie charts show the proportions of proteins with varying tissue expression patterns, from a recently published human tissue proteome map<sup>46</sup>, comparing the full set of 20,258 human proteins (**a**) with the 2,131 proteins within the identified complexes (**b**). Consistent with these observations, 91% of the protein components in the complexes were expressed in >15 tissues in data from a reference human

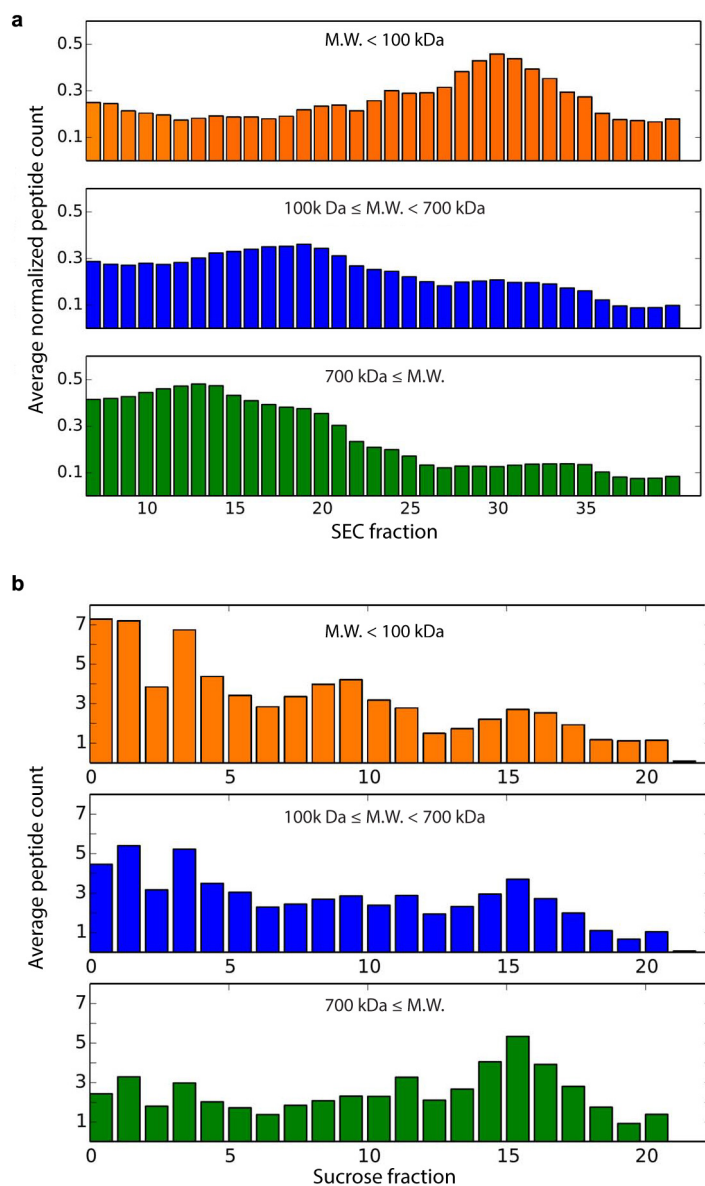
proteome<sup>23</sup>, compared to less than half (46%) of the 17,294 proteins in the overall reference set ( $Z$ -test  $P < 0.001$ ). **c, d**, The distributions of average mRNA (**c**, data from EBI accession E-MTAB-1733) and protein (**d**, data from PaxDb integrated data set, 9606-H.sapiens\_whole\_organism-integrated\_data set) abundances for all proteins identified and those within complexes. Evolutionarily old proteins (defined by OMA as described in ref. 25 and mentioned earlier) tend towards higher abundances, even for proteins in reference complexes.

46. Uhlen, M. *et al.* Tissue-based map of the human proteome. *Science* **347**, 6220 (2015).



**Extended Data Figure 6 | Additional validation data.** **a**, Confirmation of MIB2 interactions by co-immunoprecipitation. Extract (~10 mg protein) from cultured human HCT116 cells expressing Flag-tagged MIB2 or control (WT) cells was incubated with 100  $\mu$ l anti-Flag M2 resin for 4 h while gently rotating at 4 °C. After extensive washing with RIPA buffer, co-purifying proteins bound to the beads were eluted by the addition of 25  $\mu$ l Laemmli loading buffer at 95 °C. Polypeptides were separated by SDS-PAGE and immunoblotted using Flag, VPS4A, VPS4B or IST1 antibodies as indicated

(expanded gel images provided in Supplementary Information). **b**, Protein co-complex interactions reported in the CYC2008 yeast protein complex database<sup>42</sup> are reconstructed accurately from the co-fractionation data, regardless of whether the full set of co-fractionation plus external data are used to derive protein interactions ('All data', see also Fig. 4b) or if the external yeast data was specifically excluded from the analyses ('All data, excluding yeast').

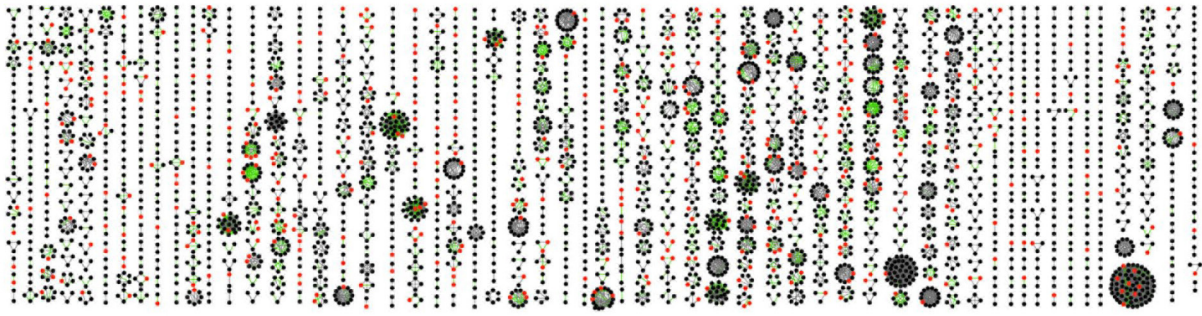


**Extended Data Figure 7 | Agreement of derived complexes' molecular weights with measurement by HPLC and density centrifugation.**

**a**, CORUM reference complexes' inferred molecular weights (MW) are consistent with their components' average cumulative size-exclusion chromatograms. The molecular weight of each complex was calculated as the sum of putative component molecular weights, assuming 1:1 stoichiometry. Data from ref. 43 were analysed as in Fig. 4c and show a similar trend as

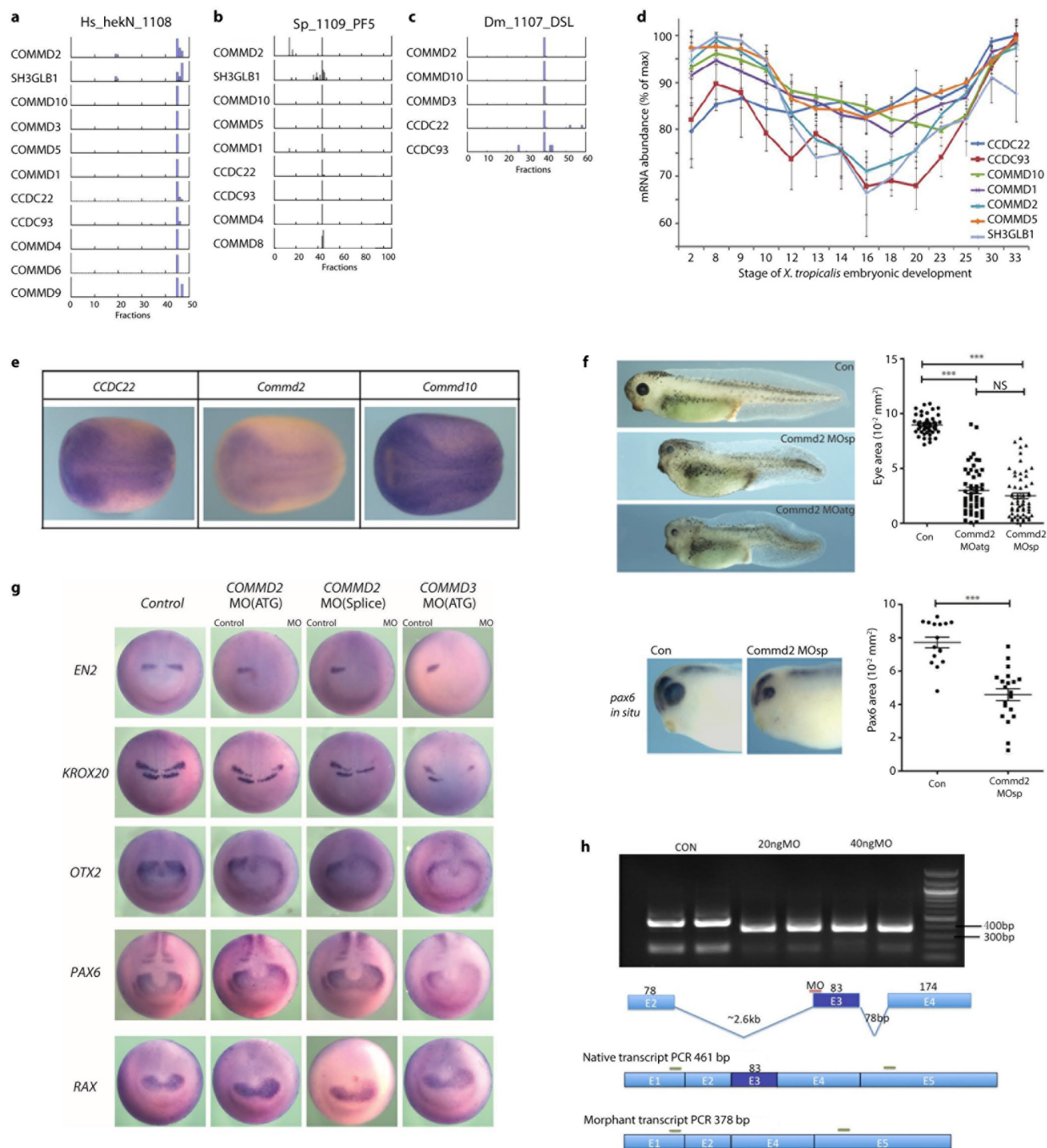
for the derived complexes. **b**, Derived complexes' inferred molecular weights are broadly consistent with their components' average cumulative ultracentrifugation profiles on a sucrose density gradient. Average profiles are plotted for *X. laevis* orthologues, based on a preparation of haemoglobin-depleted heart and liver proteins separated on a 7–47% sucrose density gradient, as described in the Supplementary Methods.





**Extended Data Figure 8 | Distribution of uncharacterized proteins and novel interactions across the 981 derived complexes.** Complexes were sorted by median age (defined by OMA). Among 2,153 unique proteins, 293 (red)

lack Gene Ontology (GO) functional annotations, while 1,756 of 7,665 co-complex interactions are novel (light green) (not listed in iRefWeb curation database).



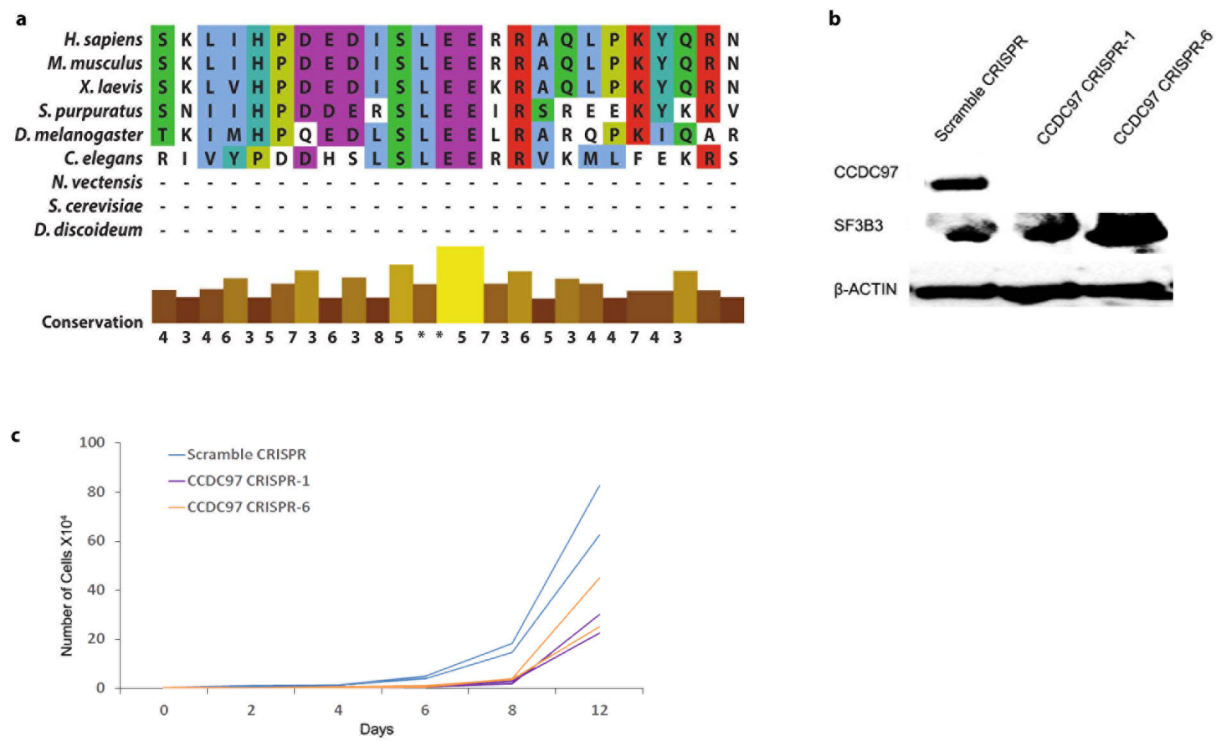
**Extended Data Figure 9 | Properties of the Commander complex.** The automatically derived 8 subunit Commander complex (Fig. 3b) was subsequently extended to 13 subunits (COMMD1 to 10, CCDC22, CCDC93, and SH3GLB1) based on combined analysis of AP/MS (Fig. 4a), size-exclusion chromatograms<sup>43</sup> (Fig. 4d), published pairwise interactions<sup>30,47,48</sup>, and analysis of elution profiles of the remaining COMM-domain-containing proteins, as shown here. Example protein elution profiles are plotted for Commander complex subunits observed from: HEK293 cell nuclear extract (a); sea urchin embryonic (5 days post-fertilization) extract (b); and fly SL2 cell nuclear extract (c); each fractionated by heparin affinity chromatography. **d**, Co-expression of Commander complex subunits during embryonic development of *X. tropicalis* (plotting mean  $\pm$  s.d. of three clutches; data from ref. 49). **e**, Messenger RNA expression patterns of Commander complex subunits in stage 15 *X. laevis* embryos. Images show coordinated spatial expression in early vertebrate embryogenesis, as measured by *in situ* hybridization (three embryos examined). **f**, Knockdown of *Commmd2* induced marked head and eye defects in developing *X. laevis*. Top, *Commmd2* antisense knockdown significantly decreased eye size, shown for stage 38 tadpoles

(from three clutches; control  $n = 47$  animals, one eye each; \*\*\* $P < 0.0001$ , two-sided Mann–Whitney test); phenotypes were consistent between translation blocking (MOatg;  $n = 60$ ) morpholino reagents, splice site blocking (MOsp;  $n = 50$ ) morpholinos, and knockdowns of interaction partner *Commmd3* (see Fig. 5a). Bottom, *Commmd2*-knockdown induced altered *Pax6* patterning in the embryonic eye (control  $n = 8$  animals, two eyes each; MO  $n = 11$ ). **g**, *Commmd2/3*-knockdown animals show altered neural patterning. Changes in stage 15 *X. laevis* embryos, measured by *in situ* hybridization (assayed in duplicates; five embryos per treatment), seen upon knockdown but not on controls: the forebrain marker *PAX6* was expanded, while the mid-brain marker *EN2* was strongly reduced. Notably, while expression of *KROX20/EGR1* in rhombomere R3 was shifted posteriorly, expression in R5 was strongly reduced or entirely absent. Panels in Fig. 5b are reproduced from this figure and are directly comparable. **h**, Confirmation of splice-blocking *Commmd2* morpholino activity. Images and schematic show the basis and results of RT–PCR and agarose gel electrophoresis obtained with the corresponding *X. laevis* knockdown tadpoles.

47. de Bie, P. *et al.* Characterization of COMMD protein–protein interactions in NF- $\kappa$ B signalling. *Biochem. J.* **398**, 63–71 (2006).
48. Phillips-Krawczak, C. A. *et al.* COMMD1 is linked to the WASH complex and regulates endosomal trafficking of the copper transporter ATP7A. *Mol. Biol. Cell* **26**, 91–103 (2015).

49. Yanai, I., Peshkin, L., Jorgensen, P. & Kirschner, M. W. Mapping gene expression in two *Xenopus* species: evolutionary constraints and developmental flexibility. *Dev. Cell* **20**, 483–496 (2011).





**Extended Data Figure 10 | Supporting data for BUB3 and CCDC97 experiments.** **a**, Sequence alignment showing conservation of ZNF207 GLEBS domain. **b**, Targeted CRISPR/Cas9-induced knockout of *CCDC97* in two independent lines of human HEK293 cells, as verified by western blotting

(expanded gel images provided in Supplementary Information). **c**, Loss of *CCDC97* impairs cell growth. Lines show growth curves of control versus knockout cell lines in two biological replicate assays.

# The mechanism of DNA replication termination in vertebrates

James M. Dewar<sup>1</sup>, Magda Budzowska<sup>1</sup> & Johannes C. Walter<sup>1,2</sup>

**Eukaryotic DNA replication terminates when replisomes from adjacent replication origins converge. Termination involves local completion of DNA synthesis, decatenation of daughter molecules and replisome disassembly. Termination has been difficult to study because termination events are generally asynchronous and sequence nonspecific. To overcome these challenges, we paused converging replisomes with a site-specific barrier in *Xenopus* egg extracts. Upon removal of the barrier, forks underwent synchronous and site-specific termination, allowing mechanistic dissection of this process. We show that DNA synthesis does not slow detectably as forks approach each other, and that leading strands pass each other unhindered before undergoing ligation to downstream lagging strands. Dissociation of the replicative CMG helicase (comprising CDC45, MCM2-7 and GINS) occurs only after the final ligation step, and is not required for completion of DNA synthesis, strongly suggesting that converging CMGs pass one another and dissociate from double-stranded DNA. This termination mechanism allows rapid completion of DNA synthesis while avoiding premature replisome disassembly.**

DNA replication occurs in three broad stages: initiation, elongation and termination. Termination occurs when converging replication forks meet and involves at least four processes, not necessarily in the following order. First, the last stretch of parental DNA between forks is unwound (dissolution) and replisomes come into contact; second, any remaining gaps in the daughter strands are filled in and nascent strands are ligated (ligation); third, double-stranded (ds)DNA intertwinings (that is, catenanes) are removed (decatenation); fourth, the replisome is disassembled. Despite decades of research on termination<sup>1</sup>, we know little about the order, mechanism and regulation of the above events, especially during eukaryotic chromosomal replication.

Termination has been most extensively studied in the mammalian DNA tumour virus SV40 (ref. 2), where converging replication forks stall during termination<sup>1,3,4</sup>. Dissolution during SV40 replication requires rotation of the entire fork to produce catenations behind the fork (pre-catenanes)<sup>5,6</sup>, which are resolved by topoisomerase (Topo) II (ref. 6), probably in a manner similar to how Topo IV functions during bacterial termination<sup>7,8</sup>. The SV40 replicative helicase, large T antigen, dissociates from chromatin before dissolution, but whether this is required for the completion of replication is unknown<sup>9,10</sup>. After dissolution, daughter strands retain gaps of ~60 nucleotides<sup>11</sup>, which are ultimately filled in by an unknown mechanism in parallel to decatenation<sup>12</sup>.

Eukaryotic termination has also been investigated. Although convergent forks accumulate at certain replication pause sites in yeast cells lacking 5'–3' DNA helicases<sup>13–15</sup>, it is unknown whether forks stall during unperturbed termination. Furthermore, Topo II is not required for dissolution in budding yeast<sup>16,17</sup> or during vertebrate termination<sup>18,19</sup>. Recent work shows that late in S phase, the eukaryotic replicative helicase CMG<sup>20–23</sup> is removed from chromatin by the ATPase p97 after ubiquitylation of MCM7 (by SCF<sup>Dia2</sup> in yeast)<sup>24,25</sup>. While one study implied that DNA replication can go to completion in the absence of CMG unloading<sup>24</sup>, another reported that tracts of unreplicated DNA remain in the absence of this process<sup>25</sup>. Given that mis-regulation of bacterial termination can readily trigger re-replication

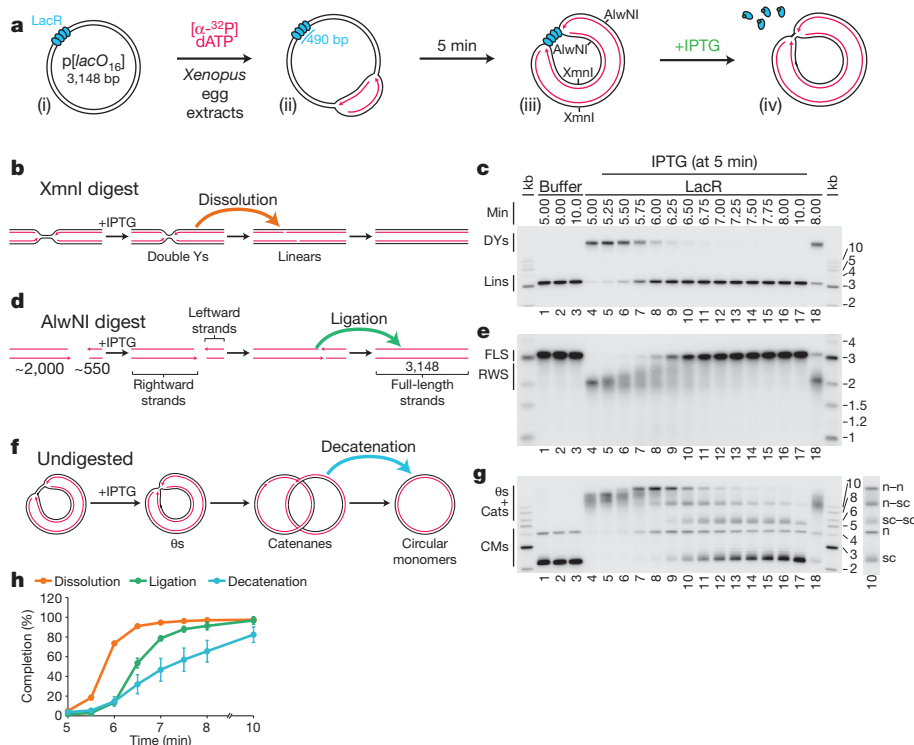
of DNA<sup>26,27</sup>, a potent driver of genomic instability in mammalian cells<sup>28</sup>, a better understanding of eukaryotic termination is essential.

Owing to stochastic origin firing<sup>29,30</sup> and variable rates of replisome progression<sup>31,32</sup>, the location and timing of eukaryotic termination is variable<sup>30,33</sup>, making this process difficult to study. Here we report that *Xenopus* egg extracts can be used to induce synchronous and localized termination events. This approach has allowed us to identify and order key events underlying vertebrate termination.

## A system to study replication termination

Our strategy was to stall forks on either side of a reversible replication fork barrier (Fig. 1a, panels i–iii), and subsequently disassemble the barrier to trigger localized and synchronous termination events (Fig. 1a, panel iv). The barrier that we employed consisted of an array of *lac* repressors (LacRs) bound to *lac* operators (*lacOs*)<sup>34,35</sup>, which can be disrupted by IPTG. We constructed p[*lacO*<sub>16</sub>], which contains 16 tandem copies of *lacO* (490 base pairs (bp)). p[*lacO*<sub>16</sub>] was incubated in nucleus-free *Xenopus* egg extract, which promotes sequence-nonspecific replication initiation on added DNA molecules, followed by a single, complete round of DNA synthesis via a mechanism that appears to reflect events in cells<sup>36</sup>. To monitor replication, radioactive [ $\alpha$ -<sup>32</sup>P]dATP was included in the reaction. When p[*lacO*<sub>16</sub>] was replicated in the absence of LacR for ~5 min and then cut with XmnI (Fig. 1a, panel iii), a single linear species representing fully replicated daughter molecules was observed (Fig. 1c, lane 1). In contrast, in the presence of LacR, a slow-mobility product appeared (Fig. 1c, lane 4) that corresponds to a double-Y structure, as shown by 2D gel electrophoresis (Extended Data Fig. 1a). To confirm that the double-Y resulted from fork stalling at the outer edges of the array, we separately monitored replication in the plasmid backbone and in the *lacO* array. In the presence of LacR, synthesis of the array was specifically delayed (Extended Data Fig. 1f). In contrast, LacR had no effect on replication of a plasmid lacking *lacO* sites (Extended Data Fig. 1e). These results indicate that replication forks stalled on both sides of the LacR array, consistent with previous findings<sup>34,35,37</sup>.

<sup>1</sup>Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts 02115, USA. <sup>2</sup>Howard Hughes Medical Institute, Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts 02115, USA.



**Figure 1 | A model system to study replication termination.** **a**, Scheme to induce site-specific termination. Key restriction sites are highlighted. **b**, Schematic of the dissolution assay. **c**, p[lacO<sub>16</sub>] was incubated in buffer or LacR, then replicated in the presence of [α-<sup>32</sup>P]dATP, before termination was induced by the addition of IPTG. To measure dissolution, radiolabelled termination intermediates were cut with XmnI, separated on a native agarose gel, and analysed by autoradiography. **d**, Schematic of the ligation assay. **e**, To measure ligation, replication intermediates were cut with AlwNI and separated on a denaturing agarose gel. **f**, Schematic of the decatenation assay. **g**, To measure decatenation, replication intermediates were separated on a native agarose gel. The additional copy of lane 10 highlights catenated termination intermediates. Cats, catenanes; CMs, circular monomers; DYs, double-Y structures; FLS, full-length strands; Lins, linears; LWS, leftward strands; n-n, nicked-nicked; n-sc, nicked-supercoiled; RWS, rightward strands; sc-sc, supercoiled-supercoiled; kb, kilobase ladder, with the size of each band (in kilobases) labelled. **h**, Multiple dissolution, ligation and decatenation assays were quantified. Means ± standard deviation (s.d.) are plotted (*n* = 4).

We next addressed whether replication forks stalled by LacR could restart. When IPTG was added to double-Y structures 5 min after replication initiation, 90% were converted to unit-sized linear plasmid molecules within a further 1.5 min (Fig. 1c, lanes 5–10 and Fig. 1h, yellow circles). In the absence of IPTG, only 21% of double-Y molecules disappeared after 3 min (Fig. 1c, lane 18). The conversion of double-Y molecules to linear species occurs when any remaining parental DNA holding daughter molecules together is unwound (Fig. 1b). This process, which we refer to as ‘dissolution’, represents a convenient means to measure the point at which converging replisomes meet. Notably, the ATR–Chk1 pathway was not activated above background levels during this procedure (data not shown).

After dissolution, nascent strands should undergo ligation. To detect the growth and ligation of nascent strands, we digested p[lacO<sub>16</sub>] with AlwNI, which cuts the plasmid once, ~550 nucleotides (nt) from the rightward edge of the array and ~2,000 nt from its leftward edge (Fig. 1a, panel iii, and Fig. 1d), and we analysed the products on a denaturing gel. Before IPTG addition, discrete species of ~2,000 nt (Fig. 1e, lane 4) and ~550 nt (Extended Data Fig. 2a, lane 4) were observed. Upon IPTG addition, both bands grew heterogeneously (Fig. 1e and Extended Data Fig. 2a). Since all leading strands were immediately extended upon IPTG addition (Extended Data Fig. 2b, c), we infer that the heterogeneity observed resulted because growth of the lagging strand was delayed until ligation of an additional Okazaki fragment. Finally, the nascent strands increased abruptly to the full length of 3,100 nt as ligation to downstream lagging strands occurred (Fig. 1e, lanes 9–13). As expected, dissolution preceded ligation, and there was an ~45 s delay between these two events (Fig. 1h).

Another important event associated with termination is decatenation of daughter molecules<sup>18</sup>. To measure this process, we analysed undigested replication products on native agarose gels (Fig. 1f, g). Before addition of IPTG, when the array had not yet been duplicated, replication products migrated as a compact smear of high-molecular-weight θ structures (Fig. 1f and Fig. 1g, lane 4). Upon addition of IPTG, most θ structures were lost within 1 min, and they were successively converted into three types of dimeric catenanes described previously<sup>5,18,38</sup>: nicked–nicked, nicked–supercoiled,

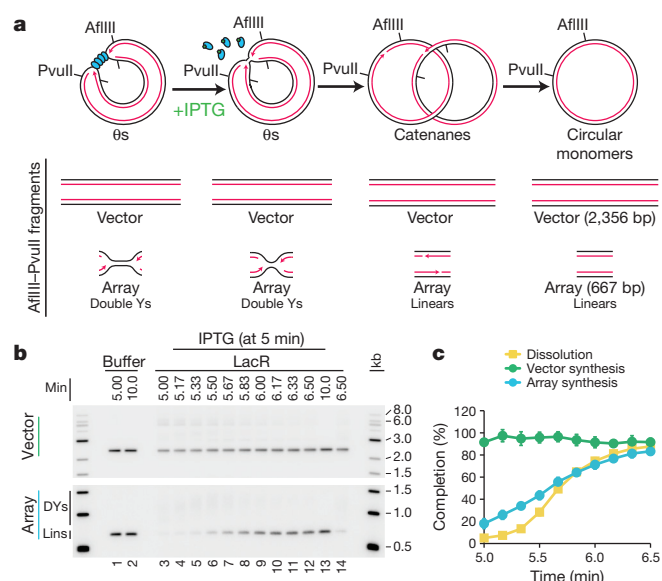
and supercoiled–supercoiled (Fig. 1g and Extended Data Fig. 3a). Nicked–nicked catenanes appeared first (Fig. 1g, lanes 7, 8), followed by nicked–supercoiled (lanes 8–10) and supercoiled–supercoiled (Fig. 1g, lanes 9–12). Supercoiling is the result of nucleosome assembly on closed circular DNA<sup>39</sup>. Finally, monomeric, supercoiled daughter molecules accumulated (sc, Fig. 1g, lane 17) dependent on Topo II (Extended Data Fig. 3b–d) as seen *in vivo*<sup>16,17</sup>. Topo II was not required for dissolution or ligation (Extended Data Fig. 3c, d), suggesting that these processes proceed independently of decatenation<sup>16,17,19</sup>. Like ligation, decatenation began ~40 s after dissolution, but progressed at a slower rate than ligation (Fig. 1h). The same intermediates were detected in the absence of LacR, but their order of appearance was not well defined (Extended Data Fig. 3e).

Our results demonstrate that a reversible replication fork barrier allows induction of a synchronous and spatially defined termination event. They also show that soon after forks meet, as measured by dissolution, daughter molecules are quickly ligated and decatenated.

### Converging replication forks do not stall

To test the proposal that replication forks slow down or stall during termination<sup>1,3,4</sup>, we quantified the rate of DNA synthesis as two replisomes converged within the *lacO* array. To minimize the loss of synchrony among replisomes after IPTG addition, we used a 365-bp array containing only 12 copies of *lacO*, which was sufficient to prevent dissolution at the 5 min time point (Extended Data Fig. 4c). We replicated p[lacO<sub>12</sub>] in the presence of LacR, added IPTG after 5 min, and examined subsequent replication within the array by cutting the plasmid with AflIII and PvuII (Fig. 2a). The rate of DNA synthesis within the array was almost perfectly linear after IPTG addition (Fig. 2b, c) even as dissolution was underway. These data suggest that converging forks do not slow significantly before they meet. A similar conclusion was reached when radiolabelled nucleotides were added at the same time as IPTG and incorporation measured only during the final stage of replication on p[lacO<sub>12</sub>] (Extended Data Fig. 5a–f) or p[lacO<sub>16</sub>] (Extended Data Fig. 5g, h). Moreover, fork rates within the *lacO* array resembled those previously reported in the same egg extracts (Extended Data Fig. 5f). These results suggest that converging replisomes do not undergo prolonged stalling.

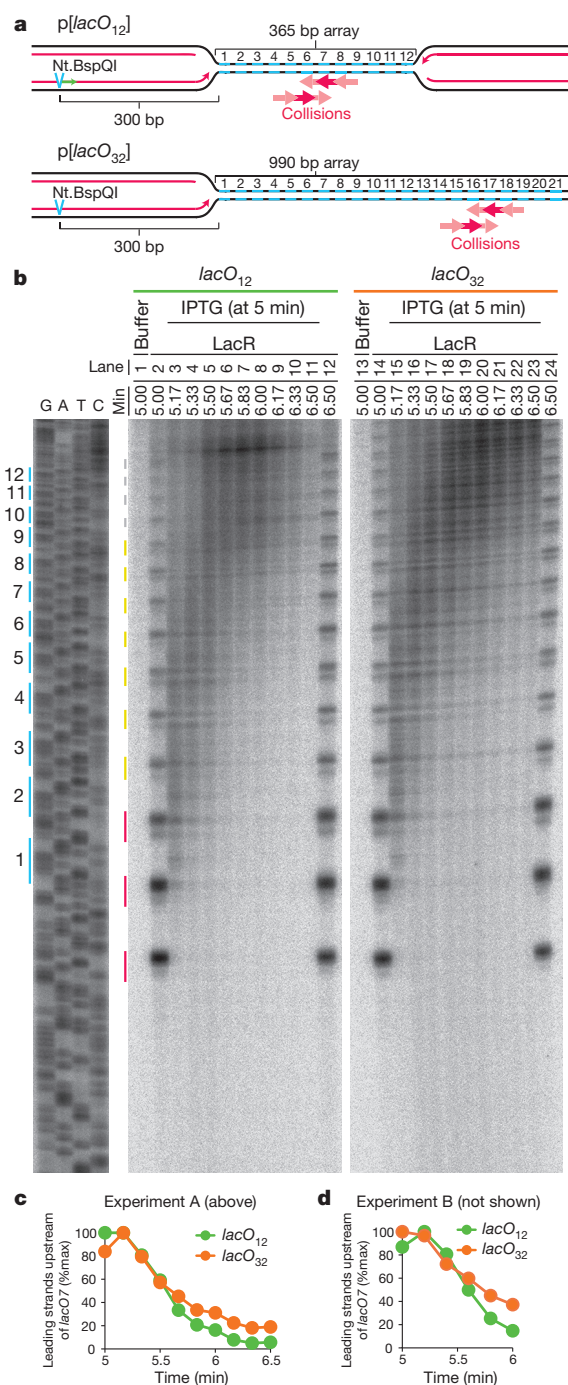




**Figure 2 | DNA synthesis does not stall during termination.** **a**, Cartoon depicting the assay for *lacO* array synthesis. **b**, LacR block-IPTG release was performed on p[*lacO*<sub>12</sub>]. To measure synthesis within the array, termination intermediates were cut with *AflIII* and *PvuII* to liberate the array fragment from the vector. Cleaved products were separated by native gel electrophoresis. Different exposures of array and vector fragments are shown (see Methods). **c**, Array synthesis, vector synthesis and dissolution were quantified. Means  $\pm$  s.d. are plotted ( $n = 3$ ). kb, kilobase ladder, with the size of each band (in kilobases) labelled.

To evaluate further whether forks slow or stall upon encounter with a converging fork, we compared progression of leading strands into arrays containing 12 or 32 copies of *lacO* (Fig. 3a), in which the rightward fork should collide with a converging fork at the 6th or 16th *lacO* repeats, respectively (Fig. 3a). If converging forks interfere with each other, the rightward leading strand should pause or stall near the 6th repeat in p[*lacO*<sub>12</sub>] but not in p[*lacO*<sub>32</sub>]. As expected, dissolution (Fig. 1b) happened much earlier on p[*lacO*<sub>12</sub>] than on p[*lacO*<sub>32</sub>] (Extended Data Fig. 6a–d). To monitor leading-strand progression into the array with near-nucleotide resolution, DNA intermediates were purified, digested with the nicking enzyme *Nt.BspQI*, which released rightward leading strands (Fig. 3a), and separated on a denaturing polyacrylamide gel (Fig. 3b). Before IPTG addition, a discrete ladder of leading strands was seen (Fig. 3b, lanes 2, 14), in which the 3' ends of leading strands stalled  $\sim 29$ –33 nt from each LacR molecule in the array. This  $\sim 30$  nt gap probably corresponds to the footprint of the CMG complex<sup>35,40</sup>. As shown in Fig. 3b (red lines) and quantified in Extended Data Fig. 6e, 78% of leading strands were stalled at the first three *lacO* sites, indicating that most replisomes were blocked at the outer edges of the array.

Upon addition of IPTG, extension of leading strands resumed immediately (Fig. 3b, lanes 3–11 and 15–23). Notably, there was no enhanced pausing near the 6th *lacO* repeat of the *lacO*<sub>12</sub> array versus the *lacO*<sub>32</sub> array. By 5.67 min, most leading strands had extended beyond the 6th *lacO* repeat within both arrays (Fig. 3b, lanes 6 and 18, and Extended Data Fig. 2c). This was also true for the leftward leading strands (Extended Data Fig. 6f, g). Furthermore, leading strands were extended beyond the 6th *lacO* repeat in the *lacO*<sub>12</sub> and *lacO*<sub>32</sub> arrays with similar kinetics (Fig. 3c, d). When leading strands were analysed on alkaline denaturing gels, we observed that all rightward and leftward leading strands passed the mid-point of the array by 6.25 min (Extended Data Fig. 2d, e), indicating that the converging leading strands were readily extended past each other. In summary, we failed to observe detectable slowing or pausing of DNA synthesis during termination, and converging leading strands passed each other unhindered, implying that converging replisomes do not pause or stall significantly.

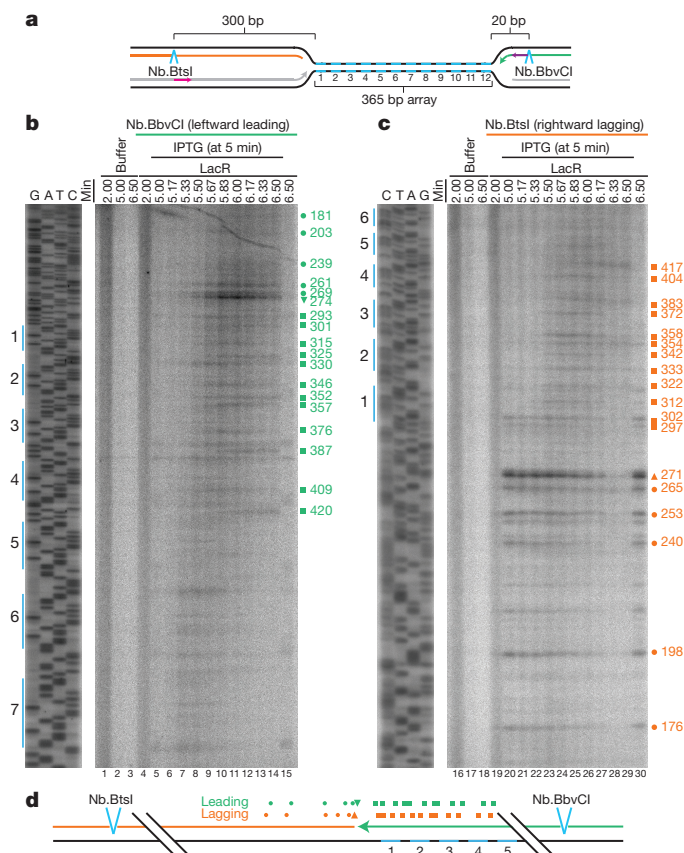


**Figure 3 | Leading strands pass each other unhindered during termination.** **a**, Schematic of rightward leading strands arrested at 12 $\times$  and 32 $\times$  *lacO* arrays (p[*lacO*<sub>12</sub>] and p[*lacO*<sub>32</sub>], respectively), and the predicted point of fork collision upon IPTG addition. **b**, LacR block-IPTG release was performed on p[*lacO*<sub>12</sub>] and p[*lacO*<sub>32</sub>]. Termination intermediates were digested with *Nt.BspQI*. Nascent strands were separated alongside a sequencing ladder (generated by primer JDO107, green arrow in **a**) on a denaturing polyacrylamide gel and visualized by autoradiography. The *lacO* sites of p[*lacO*<sub>12</sub>] are indicated in blue. Red, yellow and grey lines indicate stall products that were quantified (Extended Data Fig. 6e). **c**, Leading strands whose 3' ends were located before *lacO*<sub>7</sub> were quantified (see Methods) along with dissolution (Extended Data Fig. 6a–d). **d**, Experimental repeat of **c**.

### Lagging strand gaps are rapidly filled in

During SV40 replication termination, gaps of  $\sim 60$  nt persist after dissolution<sup>11</sup>. To determine whether the appearance of such gaps precedes the ligation step in our system, we mapped the 3' ends of

the leftward leading strands and the 5' ends of the rightward lagging strands during termination within *lacO*<sub>12</sub> (Fig. 4a). To this end, we digested DNA intermediates with Nb.BbvCI or Nb.BtsI to release leading or lagging strands, respectively (Fig. 4a), and separated them on denaturing polyacrylamide gels. After IPTG addition, we detected a prominent leading-strand product beyond the 12th *lacO* repeat (species 274 in Fig. 4b; the 3' and 5' termini of all leading and lagging strand products, respectively, are mapped relative to the Nb.BtsI site) as seen also in Fig. 3b. The 3' end of this species was located ~3 nt from the 5' end of the most abundant lagging strand product of the converging fork (271, Fig. 4c). We observed many other, less prominent, leading-strand products (181–420, Fig. 4b), most of which mapped close to corresponding lagging strand products (176–417, Fig. 4c). The results show that leading strands are generally extended to within ~3 nt of the lagging strands (Fig. 4d). It is likely that leading strands immediately abut lagging strands and that the ~3 nt gap reflects imprecise mapping of lagging strands (see Methods). In conclusion, we observed no evidence of persistent gaps between leading and lagging strands during replication termination.



**Figure 4 | Leading strands abut lagging strands of the opposing replisome during termination.** **a**, Cartoon illustrating the leading and lagging strands released by Nb.BtsI and Nb.BbvCI nicking enzymes. Primers JDO111 (purple arrow) and JDO110 (pink arrow) generated the sequencing ladders in **b** and **c**, respectively. **b**, LacR block-IPTG release was performed on p[*lacO*<sub>12</sub>]. Termination intermediates were digested with Nb.BbvCI to liberate leftward leading strands, which were separated alongside a sequencing ladder on a denaturing polyacrylamide gel and visualized by autoradiography. Prominent leading strand products are highlighted (green symbols), and their sizes, in nucleotides, measured relative to the Nb.BtsI site, are indicated. **c**, Same samples as in **b** were digested with Nb.BtsI to liberate rightward lagging strands. The size of prominent lagging strand products (orange symbols), measured relative to the Nb.BtsI site, are indicated. **d**, Schematic of the mapped leading (**b**) and lagging (**c**) strands.

## CMGs dissociate late during termination

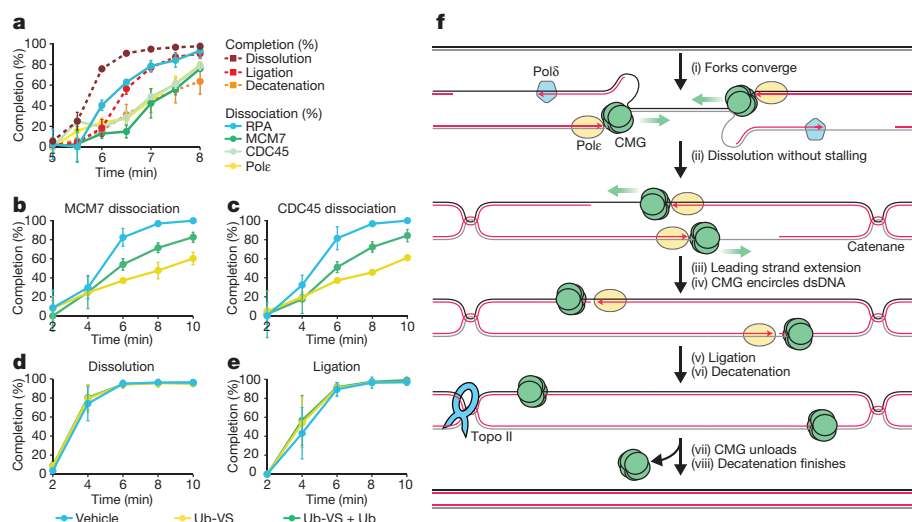
To determine when replisome components dissociate during termination, we monitored MCM7, CDC45, Polε and RPA binding to a site flanking the *lacO* array using chromatin immunoprecipitation (ChIP) (*FLK2* locus, Extended Data Fig. 7a). In parallel, we monitored dissolution, ligation and decatenation. Before IPTG addition, MCM7, CDC45, Polε and RPA were 4–8-fold enriched at the array in the presence of LacR compared to buffer (Extended Data Fig. 7b–e, 5 min time point), demonstrating that the ChIP signal reflects replisome stalling at the array. When IPTG was added at 5 min, MCM7, CDC45, RPA and Polε largely dissociated by 9 min, whereas in the absence of IPTG, they dissociated much more slowly (Extended Data Fig. 7b–e). RPA dissociation correlated well with ligation, as expected, since ligation marks the disappearance of any single-stranded (ss)DNA in the termination zone (Fig. 5a, compare red squares and blue circles). Notably, CDC45, MCM7 and Polε dissociated ~1.5 min after dissolution and ~0.5 min after RPA dissociation and ligation (Fig. 5a). A time course of ChIP at sequences adjacent to and within the array (Extended Data Fig. 7f–i) was consistent with MCM7, CDC45 and DNA Polε moving into the array and then back out after dissolution (Extended Data Fig. 7j). MCM7 and CDC45 also dissociated after dissolution during replication of plasmid DNA that lacked a *lacO* array (p[empty], Extended Data Fig. 8a, b). Although the delay between ligation and unloading of MCM7 and CDC45 was not readily detectable on this template (Extended Data Fig. 8b), this was not surprising, given the asynchrony of termination in this setting. Together, the data support a model in which CDC45 and MCM7 dissociate late in termination, long after forks meet (dissolution) and shortly after ligation.

If our model is correct, inhibiting CMG unloading should not affect dissolution or ligation. To test this, we inhibited ubiquitin signalling, which is required for chromatin dissociation of CMG<sup>24,25,41</sup>. p[empty] was replicated in extracts that were incubated with vehicle or the de-ubiquitylating enzyme inhibitor ubiquitin-vinyl-sulfone (Ub-VS), which leads to the depletion of free ubiquitin<sup>35,41</sup>, and we performed MCM7 and CDC45 ChIP. As shown in Fig. 5b, c, Ub-VS substantially delayed MCM7 and CDC45 dissociation, and this effect was partially reversed by co-addition of free ubiquitin (Fig. 5b, c). The same inhibitory effect of Ub-VS on CMG unloading was observed when plasmids were recovered from egg extract and blotted for MCM7 and CDC45 (Extended Data Fig. 8c). This analysis also confirmed previous reports<sup>24,25</sup> of MCM7 ubiquitylation during replication (Extended Data Fig. 8c). Importantly, dissolution, ligation and decatenation were not affected by Ub-VS (Fig. 5d, e and Extended Data Fig. 8f–i). We conclude that defective CMG unloading does not affect dissolution, ligation, or decatenation, strongly supporting our model that CMG unloading is a late event in replication termination.

## Discussion

We present a novel approach to induce synchronous and site-specific replication termination. Using this system, we observe no slowing or pausing of DNA synthesis as forks converge (Fig. 5f, panels i, ii). Leading strands pass each other unhindered and immediately abut downstream lagging strands before undergoing ligation (Fig. 5f, panels iii–v). CMG remains associated with DNA after dissolution, and it is unloaded only after the leading strand of one fork is ligated to the lagging strand of the opposing fork (Fig. 5f, panel vii). Catenane removal is initiated at the same time as ligation (Fig. 5f, panels v, vi). In contrast to models of termination in which replication forks stall<sup>1,3,4</sup>, our data imply that topological stress between replisomes is handled efficiently and that converging replisomes do not clash or that if they do, any remaining template DNA is immediately reeled into the stalled replisome for duplication (not shown). We previously showed that CMGs encircle the leading strand template at the replication fork<sup>42</sup>. Therefore, converging CMGs approach each other on opposite strands<sup>42,43</sup>, which helps explain how they could pass each other. If a





**Figure 5 | CMGs dissociate after dissolution and ligation.** **a**, LacR block-IPTG release was followed by MCM7, CDC45, RPA and Pole ChIP at the indicated times after IPTG addition. Dissolution, ligation and decatenation were measured in parallel. Means  $\pm$  s.d. are plotted ( $n = 3$ ). **b**, [empty] was replicated in extracts treated with vehicle, ubiquitin-vinyl sulfone (Ub-VS), or Ub-VS and free ubiquitin (Ub-VS + Ub). Dissociation of MCM7 was measured by ChIP (see Methods). Mean  $\pm$  s.d. is plotted ( $n = 3$ ). **c**, Same as **b** but CDC45 dissociation was measured. **d**, **e**, In parallel to MCM7 and CDC45 dissociation (**b**, **c**), dissolution (**d**) and ligation (**e**) were measured. Mean  $\pm$  s.d. is plotted ( $n = 3$ ). See Extended Data Fig. 8f–i for decatenation measurements and representative gels. **f**, New model of vertebrate replication termination.

fork stalls (for example, at the ribosomal DNA locus<sup>13,14</sup>), the same termination mechanism could still operate provided that the stalled fork remains stable until a converging fork arrives. We expect this to be the case, given our recent observation that a single fork stalled at a DNA interstrand cross-link does not collapse or lose its CMG complex<sup>34</sup>. We speculate that at telomeres the replisome simply runs off the chromosome end.

Our observations that CMG dissociates after the ligation step (Fig. 5a), and that ligation is not affected when CMG unloading is impaired (Fig. 5b, c, e), strongly imply that CMG is unloaded from dsDNA. We propose that when CMG reaches the 5' end of the opposing fork's lagging strand, it passes over the ssDNA–dsDNA junction and keeps moving along dsDNA (Fig. 5f), as previously observed for purified MCM2–7 and CMG *in vitro* (see refs 23, 44 but see also ref. 22). This scenario is appealing, as it would prevent CMG from interfering with ligation of the nascent strands. We propose that CMG ubiquitylation and its removal by p97 (refs 24, 25) is triggered once CMG encircles dsDNA. Such a mechanism would help to avoid inappropriate CMG unloading from active replication forks, where CMG encircles ssDNA. Our results disagree with a recent report, which concluded that inhibition of CMG unloading prevents completion of DNA synthesis<sup>25</sup>. In contrast, another report that defective CMG unloading does not prevent cell cycle progression<sup>24</sup> is consistent with our model. We recently reported that CMG can be unloaded from ssDNA when two replisomes collide with a DNA interstrand cross-link<sup>41</sup>. However, this process involves a unique, BRCA1-dependent pathway that is not employed during termination<sup>41</sup>. In conclusion, the termination mechanism described here allows rapid completion of DNA synthesis while minimizing the possibility of premature replisome disassembly.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 27 September 2014; accepted 1 July 2015.

Published online 31 August 2015.

- Levene, A. J., Kang, H. S. & Billheimer, F. E. DNA replication in SV40 infected cells. I. Analysis of replicating SV40 DNA. *J. Mol. Biol.* **50**, 549–568 (1970).
- Fanning, E. & Zhao, K. SV40 DNA replication: from the A gene to a nanomachine. *Virology* **384**, 352–359 (2009).
- Tapper, D. P. & DePamphilis, M. L. Discontinuous DNA replication: Accumulation of simian virus 40 DNA at specific stages in its replication. *J. Mol. Biol.* **120**, 401–422 (1978).
- Seidman, M. M. & Salzman, N. P. Late replicative intermediates are accumulated during simian virus 40 DNA replication *in vivo* and *in vitro*. *J. Virol.* **30**, 600–609 (1979).
- Sundin, O. & Varshavsky, A. Terminal stages of SV40 DNA replication proceed via multiply intertwined catenated dimers. *Cell* **21**, 103–114 (1980).
- Ishimi, Y., Sugawara, K., Hanaoka, F., Eki, T. & Hurwitz, J. Topoisomerase II plays an essential role as a swivelase in the late stage of SV40 chromosome replication *in vitro*. *J. Biol. Chem.* **267**, 462–466 (1992).
- Hiasa, H. & Marians, K. J. Two distinct modes of strand unlinking during theta-type DNA replication. *J. Biol. Chem.* **271**, 21529–21535 (1996).
- Espeli, O., Levine, C., Hassing, H. & Marians, K. J. Temporal regulation of topoisomerase IV activity in *E. coli*. *Mol. Cell* **11**, 189–201 (2003).
- Segawa, M., Sugano, S. & Yamaguchi, N. Association of simian virus 40 T antigen with replicating nucleoprotein complexes of simian virus 40. *J. Virol.* **35**, 320–330 (1980).
- Tack, L. C. & DePamphilis, M. L. Analysis of simian virus 40 chromosome-T-antigen complexes: T-antigen is preferentially associated with early replicating DNA intermediates. *J. Virol.* **48**, 281–295 (1983).
- Chen, M. C., Birkenmeier, E. & Salzman, N. P. Simian virus 40 DNA replication: characterization of gaps in the termination region. *J. Virol.* **17**, 614–621 (1976).
- Sundin, O. & Varshavsky, A. Arrest of segregation leads to accumulation of highly intertwined catenated dimers: dissection of the final stages of SV40 DNA replication. *Cell* **25**, 659–669 (1981).
- Ivessa, A. S., Zhou, J. Q. & Zakian, V. A. The *Saccharomyces* Pif1p DNA helicase and the highly related Rrm3p have opposite effects on replication fork progression in ribosomal DNA. *Cell* **100**, 479–489 (2000).
- Steinacher, R., Osman, F., Dalgaard, J. Z., Lorenz, A. & Whitby, M. C. The DNA helicase Pfh1 promotes fork merging at replication termination sites to ensure genome stability. *Genes Dev.* **26**, 594–602 (2012).
- Fachinetti, D. et al. Replication termination at eukaryotic chromosomes is mediated by Top2 and occurs at genomic loci containing pausing elements. *Mol. Cell* **39**, 595–605 (2010).
- DiNardo, S., Voelkel, K. & Sternglanz, R. DNA topoisomerase II mutant of *Saccharomyces cerevisiae*: topoisomerase II is required for segregation of daughter molecules at the termination of DNA replication. *Proc. Natl Acad. Sci. USA* **81**, 2616–2620 (1984).
- Baxter, J. & Diffley, J. F. Topoisomerase II inactivation prevents the completion of DNA replication in budding yeast. *Mol. Cell* **30**, 790–802 (2008).
- Lucas, I., Germe, T., Chevrier-Miller, M. & Hyrien, O. Topoisomerase II can unlink replicating DNA by precatenane removal. *EMBO J.* **20**, 6509–6519 (2001).
- Gaggioli, V., Le Viet, B., Germe, T. & Hyrien, O. DNA topoisomerase II controls replication origin cluster licensing and firing time in *Xenopus* egg extracts. *Nucleic Acids Res.* **41**, 7313–7331 (2013).
- Ivessa, I., Petojevic, T., Pesavento, J. J. & Botchan, M. R. Activation of the MCM2–7 helicase by association with Cdc45 and GINS proteins. *Mol. Cell* **37**, 247–258 (2010).
- Pacek, M., Tutter, A. V., Kubota, Y., Takisawa, H. & Walter, J. C. Localization of MCM2–7, Cdc45, and GINS to the site of DNA unwinding during eukaryotic DNA replication. *Mol. Cell* **21**, 581–587 (2006).
- Moyer, S. E., Lewis, P. W. & Botchan, M. R. Isolation of the Cdc45/Mcm2–7/GINS (CMG) complex, a candidate for the eukaryotic DNA replication fork helicase. *Proc. Natl Acad. Sci. USA* **103**, 10236–10241 (2006).
- Kang, Y. H., Galal, W. C., Farina, A., Tappin, I. & Hurwitz, J. Properties of the human Cdc45/Mcm2–7/GINS helicase complex and its action with DNA polymerase  $\epsilon$  in rolling circle DNA synthesis. *Proc. Natl Acad. Sci. USA* **109**, 6042–6047 (2012).
- Maric, M., Maculins, T., De Piccoli, G. & Labib, K. Cdc48 and a ubiquitin ligase drive disassembly of the CMG helicase at the end of DNA replication. *Science* **346**, 1253596–1253596 (2014).
- Priego Moreno, S., Bailey, R., Campion, N., Herron, S. & Gambus, A. Polyubiquitylation drives replisome disassembly at the termination of DNA replication. *Science* **346**, 477–481 (2014).
- Hiasa, H. & Marians, K. J. Tus prevents overreplication of oriC plasmid DNA. *J. Biol. Chem.* **269**, 26959–26968 (1994).



27. Rudolph, C. J., Upton, A. L., Stockum, A., Nieduszynski, C. A. & Lloyd, R. G. Avoiding chromosome pathology when replication forks collide. *Nature* **500**, 608–611 (2013).
28. Hook, S. S., Lin, J. J. & Dutta, A. Mechanisms to control rereplication and implications for cancer. *Curr. Opin. Cell Biol.* **19**, 663–671 (2007).
29. Czajkowsky, D. M., Liu, J., Hamlin, J. L. & Shao, Z. DNA combing reveals intrinsic temporal disorder in the replication of yeast chromosome VI. *J. Mol. Biol.* **375**, 12–19 (2008).
30. McGuffee, S. R., Smith, D. J. & Whitehouse, I. Quantitative, genome-wide analysis of eukaryotic replication initiation and termination. *Mol. Cell* **50**, 123–135 (2013).
31. Yardimci, H., Loveland, A. B., Habuchi, S., van Oijen, A. M. & Walter, J. C. Uncoupling of sister replisomes during eukaryotic DNA replication. *Mol. Cell* **40**, 834–840 (2010).
32. Loveland, A. B., Habuchi, S., Walter, J. C. & van Oijen, A. M. A general approach to break the concentration barrier in single-molecule imaging. *Nature Methods* **9**, 987–992 (2012).
33. Santamaria, D. *et al.* Bi-directional replication and random termination. *Nucleic Acids Res.* **28**, 2099–2107 (2000).
34. Zhang, J. *et al.* DNA interstrand cross-link repair requires replication fork convergence. *Nature Struct. Mol. Biol.* **22**, 242–247 (2015).
35. Duxin, J. P. P., Dewar, J. M. M., Yardimci, H. & Walter, J. C. C. Replication-coupled repair of a DNA-protein crosslink. *Cell* **159**, 346–357 (2014).
36. Walter, J., Sun, L. & Newport, J. Regulated chromosomal DNA replication in the absence of a nucleus. *Mol. Cell* **1**, 519–529 (1998).
37. Sofueva, S. *et al.* Ultrafine anaphase bridges, broken DNA and illegitimate recombination induced by a replication fork barrier. *Nucleic Acids Res.* **39**, 6568–6584 (2011).
38. Charbin, A., Bouchoux, C. & Uhlmann, F. Condensin aids sister chromatid decatenation by topoisomerase II. *Nucleic Acids Res.* **42**, 340–348 (2014).
39. Laskey, R. A., Mills, A. D. & Morris, N. R. Assembly of SV40 chromatin in a cell-free system from *Xenopus* eggs. *Cell* **10**, 237–243 (1977).
40. Räschle, M. *et al.* Mechanism of replication-coupled DNA interstrand crosslink repair. *Cell* **134**, 969–980 (2008).
41. Long, D. T. T., Joukov, V., Budzowska, M. & Walter, J. C. C. BRCA1 promotes unloading of the CMG helicase from a stalled DNA replication fork. *Mol. Cell* **56**, 174–185 (2014).
42. Fu, Y. V. *et al.* Selective bypass of a lagging strand roadblock by the eukaryotic replicative DNA helicase. *Cell* **146**, 931–941 (2011).
43. Costa, A. *et al.* The structural basis for MCM2–7 helicase activation by GINS and Cdc45. *Nature Struct. Mol. Biol.* **18**, 471–477 (2011).
44. Kaplan, D. L., Davey, M. J. & O'Donnell, M. Mcm4,6,7 uses a 'pump in ring' mechanism to unwind DNA by steric exclusion and actively translocate along a duplex. *J. Biol. Chem.* **278**, 49171–49182 (2003).

**Acknowledgements** We thank C. Richardson and members of the Walter laboratory for feedback on the manuscript. We thank K. J. Mariani and J. T. Yeeles for plasmids and the LacI purification protocol. J.C.W. was supported by NIH grants GM62267 and GM80676. J.C.W. is an investigator of the Howard Hughes Medical Institute.

**Author Contributions** J.M.D. and J.C.W. designed the experiments. J.M.D. performed the experiments. M.B. developed methodologies for plasmid pull downs and HIS<sub>6</sub>-Ub immunoprecipitations. J.M.D. and J.C.W. interpreted the data and wrote the paper.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.C.W. ([johannes\\_walter@hms.harvard.edu](mailto:johannes_walter@hms.harvard.edu)).

## METHODS

No statistical methods were used to predetermine sample size.

**Protein purification.** Biotinylated LacR was purified using a protocol adapted from Kenneth Marian's laboratory (personal communication). The LacR open reading frame was fused to a C-terminal AviTag (Avidity, Denver, CO) and expressed from pET11a (pET11a[LacR-Avi]). To biotinylate the AviTag on LacR-Avi, biotin ligase was co-expressed from pBirAcm (Avidity, Denver, CO). To this end, pET11a[LacR-Avi] and pBirAcm were co-transformed into T7 Express cells (New England Biolabs) and grown in the presence of ampicillin ( $100 \mu\text{g ml}^{-1}$ ) and chloramphenicol ( $17 \mu\text{g ml}^{-1}$ ). Expression of LacR-Avi and the biotin ligase was induced by addition of IPTG to a final concentration of 1 mM. Cultures were supplemented with  $50 \mu\text{M}$  biotin (Research Organics, Cleveland, OH) to ensure efficient biotinylation of LacR-Avi.

Cell pellets were resuspended in lysis buffer (50 mM Tris-HCl, pH 7.5, 5 mM EDTA, 100 mM NaCl, 1 mM DTT, 10% sucrose (w/v), Complete protease inhibitor (Roche, Nutley, NJ)). The cells were lysed at room temperature in the presence of  $0.2 \text{ mg ml}^{-1}$  lysozyme and 0.1% Brij 58. The insoluble, chromatin-containing fraction was isolated by centrifugation at  $4^\circ\text{C}$ . Chromatin-bound LacR was then released by sonication (in 50 mM Tris-HCl, pH 7.5, 5 mM EDTA, 1 M NaCl, 1 mM DTT, Complete protease inhibitor, 30 mM IPTG). DNA was removed from the soluble fraction by addition of polymin P (final concentration 1%), LacR was precipitated by addition of ammonium sulfate (final concentration 37%). The precipitate was dissolved in wash buffer (50 mM Tris-HCl, pH 7.5, 1 mM EDTA, 2.5 M NaCl, 1 mM DTT, Complete protease inhibitor) and then applied to a column of SoftLink avidin resin (Promega, Madison, WI). LacR was eluted (in 50 mM Tris-HCl, pH 7.5, 1 mM EDTA, 100 mM NaCl, 1 mM DTT, 5 mM biotin) and dialysed overnight (against 50 mM Tris-HCl, pH 7.5, 1 mM EDTA, 150 mM NaCl, 1 mM DTT, 38% glycerol (v/v)). Purified LacR was frozen in liquid nitrogen and stored at  $-80^\circ\text{C}$ . A more detailed purification protocol is available on request.

Cyclin A was purified as described previously<sup>45</sup>.

**Plasmid construction and preparation.** pJD82 (Extended Data Table 1) was created by replacing the SacI-KpnI fragment of pBlueScript II KS- with the sequence: GAGCTCTCACACCTACAAGGGATGTACATCAATTGTGAGCG GATAACAATTGTTAGGGAGGAATTGTGAGCGGATAACAATTGGAGT TGATAATTGTGAGCGGATAACAATTGGCTTCAACGTAATTGTGAGCGG ATAACAATTTCGGTACGAATGTGCCGAACCTATGTTACC. This contains four tandem repeats of the *lac* operator sequence (AATTGTGAGCGGATAACAATT) interspersed by an average of 10–11 bp of random sequence (average 10.33 bp). Additional tandem repeats of the BsiWI-BsrGI fragment were then cloned into pJD82, and subsequently derived vectors, to generate arrays of 8, 12, 16, 32 and 48 *lacO* repeats (Extended Data Table 1). Recognition sites for nicking enzymes were introduced by QuickChange mutagenesis (Agilent Technologies, Santa Clara, CA) according to the manufacturer's guidelines.

To propagate *lacO* plasmid DNA, plasmids were transformed into DH5 $\alpha$  cells and grown for a minimal number of passages in the presence of 2 mM IPTG. DNA was prepared using the QIAprep spin kit (Qiagen, Valencia, CA). To eliminate preparations containing genetic rearrangements (typically ~25%), each preparation was separated by electrophoresis on a 0.8% agarose gel and visualized by ethidium bromide staining. Preparations that were free of rearranged plasmids were then verified by sequencing (Genewiz, Cambridge, MA).

**Xenopus egg extracts and DNA replication.** *Xenopus* egg extracts were prepared from *Xenopus laevis* wild-type males and females 2–5 years of age, as approved by the Harvard Medical School Institutional Animal Care and Use Committee (IACUC) and as described previously<sup>46</sup>. For DNA replication, 1 volume of 'licensing mix' was prepared by adding plasmid DNA to High Speed Supernatant (HSS) of egg cytoplasm to a final concentration of  $7.5\text{--}15 \text{ ng } \mu\text{l}^{-1}$ . Licensing mix was incubated for 30 min at room temperature, leading to the formation of pre-replication complexes (pre-RCs). Next, licensing mix was supplemented with 0.1 volumes of cyclin A to a final volume of 576 nM and incubated a further 10 min at room temperature, as previously described<sup>45</sup>. Cyclin A treatment was performed to achieve highly synchronous DNA replication (Extended Data Fig. 9). Finally, 1.9 volumes of nucleoplasmic extract (NPE) was added to initiate Cdk2-dependent replication at pre-RCs. In all figures, '0 minutes' represents the time 30 s after NPE addition. To radiolabel DNA, NPE was supplemented with [ $\alpha\text{-}^{32}\text{P}$ ]dATP. Reactions were stopped with 10 volumes Stop Solution (0.5% SDS, 25 mM EDTA, 50 mM Tris-HCl pH 7.5). DNA in Stop Solution was treated with RNase A ( $190 \text{ ng } \mu\text{l}^{-1}$  final concentration) then Proteinase K ( $909 \text{ ng } \mu\text{l}^{-1}$  final concentration) before either direct analysis by gel electrophoresis or purification of DNA as described previously<sup>40</sup>. For Ub-VS experiments, Ub-VS (Boston Biochem) was added to final concentration of  $20 \mu\text{M}$ , to HSS 5 min before addition of plasmid DNA (HSS) and to NPE 5 min before addition of HSS, with or

without  $120 \mu\text{M}$  ubiquitin (Boston Biochem). Unless otherwise stated in the figure legend, all experiments were performed at least twice and a representative result is shown. Replicate samples were collected from independently assembled replication reactions, and therefore represent biological replicates.

**Immunodepletions.** To deplete Topo II- $\alpha$  from *Xenopus* egg extracts one volume of Protein A Sepharose Fast Flow (PAS) (GE Healthcare) was incubated with 4.5 volumes of affinity purified, anti-Topo II- $\alpha$  antibody raised against the C-terminal 20 residues ( $1 \text{ mg ml}^{-1}$ ). For mock depletion, an equivalent quantity of nonspecific IgGs was used. Five volumes of pre-cleared HSS or NPE was then mixed with one volume of the antibody-bound sepharose and incubated for 45 min at  $4^\circ\text{C}$ , and for the NPE this was repeated once. Depleted extracts were collected and used immediately for DNA replication.

**Induction of termination.** To monitor termination, 0.05 volumes of plasmid DNA ( $150\text{--}300 \text{ ng } \mu\text{l}^{-1}$ ) was incubated with 0.1 volumes LacR ( $54 \mu\text{M}$ ) or dialysis buffer for at least 90 min at room temperature to allow formation of LacR arrays on the DNA. Licensing mix was prepared by adding 0.85 volumes of HSS, and DNA was replicated as described above. To induce termination, 0.06 volumes of IPTG was added (to a final concentration of 10 mM) at the time indicated (typically 5 min), which triggered dissociation of *lacO*-bound LacR. To accurately withdraw samples at the times indicated, reactions composed of the same Licensing Mix and NPE were staggered, where necessary.

**2D gel electrophoresis.** 2D gels were performed as described<sup>41</sup>. Briefly, purified DNA was digested with XmnI (New England BioLabs) and then separated by native-native 2D gel electrophoresis. Samples were separated in the first dimension on a 0.4% agarose gel at 0.75 volts ( $\text{V cm}^{-1}$ ) for approximately 40 h at room temperature. The gel was stained with  $0.3 \mu\text{g ml}^{-1}$  ethidium bromide, allowing the 2–8 kb size range to be excised. A second dimension gel containing 1% agarose and  $0.3 \mu\text{g ml}^{-1}$  ethidium bromide was cast over the gel slice from the first dimension. DNA was separated on the second dimension at  $4.5 \text{ V cm}^{-1}$  for 12 h at  $4^\circ\text{C}$ .

**Termination assays.** To monitor dissolution,  $0.25\text{--}1.0 \text{ ng } \mu\text{l}^{-1}$  of purified DNA was incubated in CutSmart Buffer with 0.4 units  $\mu\text{l}^{-1}$  of XmnI (New England BioLabs) at  $37^\circ\text{C}$  for 1 h. Digested products were separated on a 1.2% agarose gel at  $4 \text{ V cm}^{-1}$  and detected by autoradiography. Dissolution (%) was calculated as the percentage of total signal in each lane present in the linear products of digestion (Lins, Fig. 1c).

To monitor ligation,  $0.25\text{--}1.0 \text{ ng } \mu\text{l}^{-1}$  of purified DNA was incubated in CutSmart buffer with 0.2 units  $\mu\text{l}^{-1}$  of AlwNI (New England BioLabs) at  $37^\circ\text{C}$  for 1 h. Digests were terminated by addition of EDTA to 30 mM, then products were separated on a 1.5% denaturing alkaline agarose gel at  $1.5 \text{ V cm}^{-1}$  and detected by autoradiography. The percentage of total signal in each lane present in the full-length strands was measured (FLS, Fig. 1e). During electrophoresis, partial hydrolysis caused signal from the FLS to smear down. To correct for this, a fully ligated plasmid was cleaved and analysed on the same gel. The percentage of signal in FLS band of the fully ligated plasmid was measured (FLS<sup>FL</sup>) and used to correct signal in the other lanes to yield an accurate measure of ligation. Ligation (%) was calculated as  $\text{FLS}/\text{FLS}^{\text{FL}} \times 100$ .

To monitor decatenation,  $0.25\text{--}1.0 \text{ ng } \mu\text{l}^{-1}$  of purified DNA was separated on a 0.8% agarose gel at  $4 \text{ V cm}^{-1}$  and detected by autoradiography. Decatenation (%) was measured as the percentage of total signal in each lane present in circular monomers (CMs, Fig. 1g).

To monitor DNA synthesis within a *lacO* array (Fig. 2),  $0.25\text{--}1.0 \text{ ng } \mu\text{l}^{-1}$  of purified DNA was incubated in buffer 3.1 with 0.2 units  $\mu\text{l}^{-1}$  PvuII and 0.2 units  $\mu\text{l}^{-1}$  AflIII (New England BioLabs) at  $37^\circ\text{C}$  for 1 h. Digested products were separated on a 1.2% agarose gel at  $4 \text{ V cm}^{-1}$  and detected by autoradiography. To measure array synthesis (SYN<sup>ARY</sup>), the 0.5–1.5-kb region of each lane was quantified (Lins and DYs, Fig. 2b). To measure vector synthesis (SYN<sup>VEC</sup>), the 2–6 kb region of each lane was quantified, which included the ~3.0 and ~6.0 bands that arose when one, or both, lagging strands did not cut, respectively. Total signal in each lane (SYN<sup>TOT</sup>) was also measured. To correct for differences in efficiency of DNA extraction, total lane signal was also measured in a set of unprocessed samples (SYN<sup>UN</sup>), which were separated and detected in parallel. Array synthesis (%) was calculated as  $\text{SYN}^{\text{UN}}/\text{SYN}^{\text{TOT}} \times \text{SYN}^{\text{ARY}}$ , vector synthesis was calculated as  $\text{SYN}^{\text{UN}}/\text{SYN}^{\text{TOT}} \times \text{SYN}^{\text{VEC}}$  and in both cases the 10 min time point was assigned a value of 100%. The same approach was also used to quantify synthesis of the 294/794 bp fragments (quantified in the same manner as the array) and the 2,354 bp fragments (quantified in the same manner as the vector fragments) in Extended Data Fig. 1. In Fig. 2 and Extended Data Fig. 1, a longer exposure of the array fragment is shown because it is less intense than the vector fragment.

To analyse topoisomers (Extended Data Fig. 3d),  $0.25 \text{ ng } \mu\text{l}^{-1}$  of radiolabelled DNA was incubated in 1 $\times$  buffer A and 1 $\times$  buffer B (Topogen) with  $0.2 \text{ U } \mu\text{l}^{-1}$

Human Topo II- $\alpha$  (Topogen) at 37 °C for 15 min, or in CutSmart buffer with 0.4 U  $\mu\text{l}^{-1}$  XmnI or 0.04 U  $\mu\text{l}^{-1}$  Nt.BspQI (New England Biolabs) for 1 h.

**Nascent strand analysis.** To nick rightward leading strands, 1–2 ng  $\mu\text{l}^{-1}$  of purified DNA was incubated in buffer 3.1 with 0.4 units  $\mu\text{l}^{-1}$  Nt.BspQI (New England Biolabs) at 37 °C for 1 h. To nick leftward leading strands, 1–2 ng  $\mu\text{l}^{-1}$  of purified DNA was incubated in CutSmart buffer with 0.04 units  $\mu\text{l}^{-1}$  Nb.BsrDI (New England Biolabs) at 65 °C for 1 h. To nick rightward leading strands closer to the *lacO* array, 1–2 ng  $\mu\text{l}^{-1}$  of purified DNA was incubated in CutSmart buffer with 0.04 units  $\mu\text{l}^{-1}$  Nb.BbvCI (New England Biolabs) at 37 °C for 1 h. To nick leftward lagging strands, 1–2 ng  $\mu\text{l}^{-1}$  of purified DNA was incubated in buffer 3.1 with 0.04 units  $\mu\text{l}^{-1}$  Nb.BtsI (New England Biolabs) at 37 °C for 1 h. In all cases, nicking reactions were stopped by the addition of 0.5 volumes of Stop solution B (95% formamide, 20 mM EDTA, 0.05% bromophenol blue, 0.05% xylene cyanol FF).

Nicked DNA (1.5–2  $\mu\text{l}$  sample) was separated on a 42-cm-long 4% or 5% polyacrylamide sequencing gel using Model S2 sequencing gel apparatus (Apogee Electrophoresis, Baltimore, MD) according to the manufacturer's guidelines. To maximize the range of nascent products that could be resolved, gels were cast with a thickness gradient of 0.4 to 1.2 mm, beginning to end, to establish an electrical field gradient during electrophoresis. Sequencing gels were prepared with Rapidgel-XL in 0.8 $\times$  GTG buffer (USB Corporation, Cleveland). Sequencing ladders were generated using the Cycle Sequencing Kit (USB Corporation, Cleveland) with primers JDO107, JDO109, JDO110, JDO111 (Extended Data Table 1) and pJD150 (Extended Data Table 1) as template DNA.

Mapping and quantification of the nascent strands in Figs 3 and 4 was performed as follows. Nascent leading and lagging strands were mapped using the sequencing ladders generated by the primers indicated in Fig. 3a and Fig. 4a (see Extended Data Table 1 for sequences). Slight discrepancies may exist between mapped and actual lagging strand product sizes (Fig. 4c) since the sequencing ladder (generated by JDO110 Fig. 4a) is complementary to the lagging strands. A fraction of lagging strand products 176–302 were not extended upon IPTG addition, probably because they were reached by the rightward leading strand first. Lagging strand products 312–417 appeared *de novo* after IPTG addition, and therefore represent growing lagging strands of the leftward fork. To quantify leading strand progression (Fig. 3c, d), leading strands whose 3' ends were located before *lacO7* (in Fig. 3b and data not shown) were quantified, and peak signal was assigned a value of 100 (%Max).

**ChIP and quantitative PCR.** ChIP and quantitative PCR (qPCR) were performed essentially as described<sup>41</sup>. Chromatin was withdrawn and crosslinked in the presence of 1% formaldehyde for 10 min at room temperature. Crosslinking was then quenched by the addition of 0.1 volumes glycine (1.25 M) for 10 min. Samples were then spun through Bio-Spin P-6 Gel (containing Tris Buffer, Bio-Rad) to remove salts and small molecules, before being stored in 10 volumes of sonication buffer (20 mM Tris pH 7.5, 150 mM NaCl, 2 mM EDTA, 1% IGEPAL CA-630 (v/v), 2 mM PMSF, 5  $\mu\text{g} \mu\text{l}^{-1}$  aprotinin, 5  $\mu\text{g} \mu\text{l}^{-1}$  leupeptin). Samples were then sonicated to shear chromatin into approximately 250 bp fragments.

The antibodies used were described previously<sup>21,41</sup>. Antibodies were incubated with chromatin overnight at 4 °C, then immunoprecipitated by addition of Protein A-Sepharose Fast Flow beads (GE Healthcare) for 2 h at room temperature. Beads were washed sequentially with sonication buffer, high salt buffer (sonication buffer supplemented with 500 mM NaCl and 100 mM KCl), wash buffer (10 mM Tris pH 7.5, 0.25 M LiCl, 1 mM EDTA, 0.5% NP-40 (v/v), 0.5% SDS (w/v) and TE (10 mM Tris pH 7.5, 1 mM EDTA), before being eluted into elution buffer (50 mM Tris pH 7.5, 10 mM EDTA, 1% SDS) at 65 °C for 20 min. Eluted chromatin, and input samples, were treated with RNase for 30 min at 37 °C. Finally, proteins were degraded by addition of NaCl (250 mM final) and treatment with Pronase (2  $\mu\text{g} \mu\text{l}^{-1}$  final) at 42 °C for 6 h. DNA-peptide crosslinks were reversed by treatment at 70 °C for a further 9 h. DNA was subsequently phenol:chloroform extracted and ethanol precipitated. The absolute amount of DNA recovered from the immunoprecipitated and input samples was measured by quantitative PCR (qPCR) relative to a standard curve. The qPCR primers used are listed in Extended Data Table 1. Binding was measured as the percentage recovery of immunoprecipitated DNA, relative to the input ( $\text{EXP}^{\text{REC}}$ ).

To minimize error in the ChIP process, an internal control was built into all experiments. *Xenopus* egg extracts were used to separately replicate a different

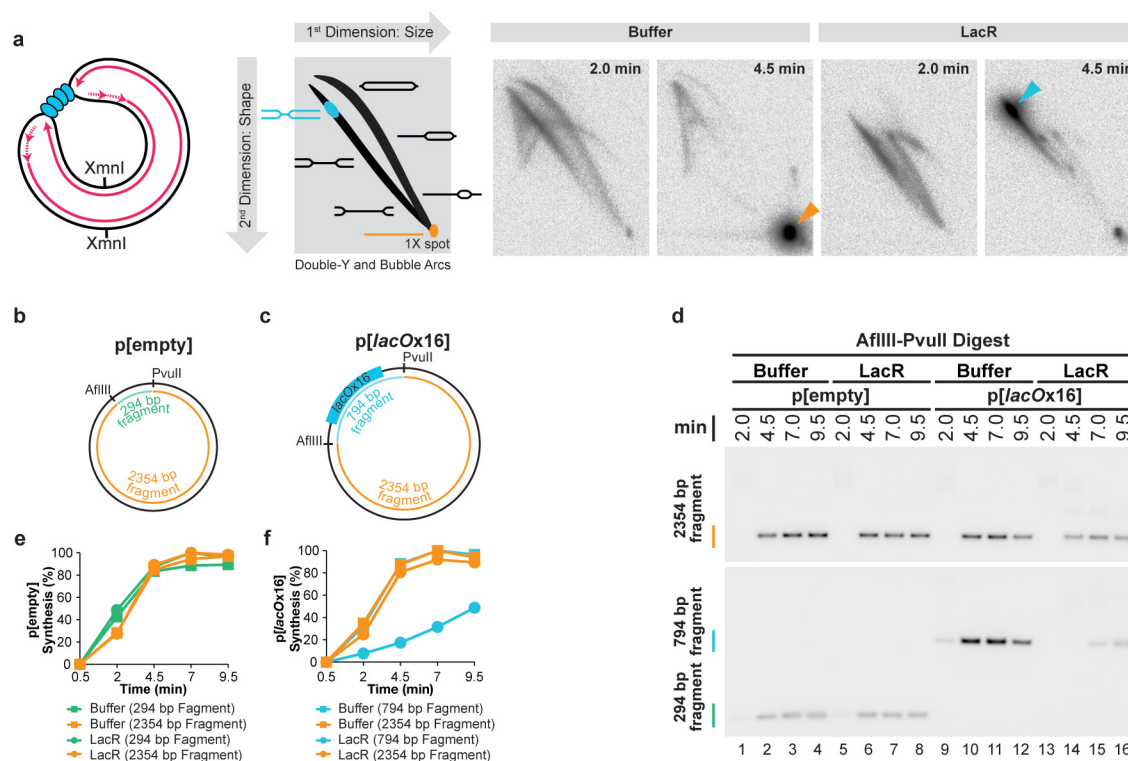
plasmid, pQUANT (see Extended Data Table 1 for sequences). Mid-way through replication, pQUANT was crosslinked, quenched and spun through Bio-Spin P-6 gel (as above) to yield a single pool of heterologous chromatin that was bound by replication proteins. An equal amount of pQUANT chromatin was added to all experimental chromatin samples before sonication, and this was carried through the entire ChIP procedure. For each set of immunoprecipitations, the recovery of pQUANT ( $\text{QNT}^{\text{REC}}$ ) should be identical between samples. To correct for technical variation in any set of immunoprecipitations, average pQUANT recovery was calculated ( $\text{QNT}^{\text{AVG}}$ ) and normalized recovery (%) was calculated as  $\text{EXP}^{\text{REC}} \times \text{QNT}^{\text{AVG}} / \text{QNT}^{\text{REC}}$ . This ensured that the only sources of technical variation were the crosslinking process and the qPCR. To maximize the reliability of the qPCR, these measurements were performed in triplicate and the median value was used. Where three ChIP experiments were combined and plotted as mean  $\pm$  s.d. (Fig. 5a–c and Extended Data Figs 7f–i and 8j) it was necessary to normalize the data to correct for differences in absolute IP efficiency between experiments. For each protein measured by ChIP, mean recovery across all loci in all samples ( $\text{mean}^{\text{all}}$ ) was calculated for each experiment ( $\text{mean}^{\text{all1}}$ ,  $\text{mean}^{\text{all2}}$  and  $\text{mean}^{\text{all3}}$ ) and used to generate a correction factor for each experiment (for example, for experiment 1 the correction factor is  $[(\text{mean}^{\text{all1}} + \text{mean}^{\text{all2}} + \text{mean}^{\text{all3}})/3]/\text{mean}^{\text{all1}}$ ). To measure dissociation (Fig. 5a–c), recovery of the *FLK2* locus was measured (shown in Extended Data Figs 7f–i and 8j) and peak signal was assigned a value of '0', while background signal (measured at 4 or 5 min for Fig. 5a, or 10 min for Fig. 5b, c) was assigned a value of 100. The experiments shown in Fig. 5a and Extended Data Fig. 7f–i were repeated three times, once with p[*lacOx12*] and twice with p[*lacOx16*].

**Plasmid pull downs.** Plasmid pull downs were performed essentially as described<sup>47</sup>, with the following exceptions. Beads were resuspended in buffer supplemented with 4% DMSO and 100  $\mu\text{M}$  NMS-873<sup>48</sup> to block further CMG unloading once the samples were withdrawn<sup>25</sup>. Plasmid-associated proteins from 40–80 ng of plasmid were isolated, and a quarter of the sample was analysed by western blotting using previously described antibodies against CDC45, MCM7 and PCNA<sup>41</sup>.

**HIS<sub>6</sub>-Ub immunoprecipitations.** Ni-NTA Superflow Resin (Qiagen) was washed three times with Urea buffer (10 mM imidazole, 0.2% NP-40, 8 M urea, 500 mM NaH<sub>2</sub>PO<sub>4</sub>, 50 mM Tris HCl, pH 8.0). For each immunoprecipitation, 10  $\mu\text{l}$  of resin was added per tube, and resuspended to 191  $\mu\text{l}$  in Urea buffer. Extracts were supplemented with 100  $\mu\text{M}$  of HIS<sub>6</sub>-ubiquitin (Boston Biochem) and replication was carried out as described above. At the indicated time, 9  $\mu\text{l}$  of extract was mixed with the bead mix and samples were incubated for 1 h at room temperature, with end-over-end rotation. Resin was washed three times with urea buffer. All residual buffer was removed, and resin was boiled for 5 min in 30  $\mu\text{l}$  sample buffer (125 mM Tris-HCl pH 6.8, 20% glycerol, 6.1% SDS, 0.01% bromophenol blue, 10%  $\beta$ -mercaptoethanol). 30  $\mu\text{l}$  of 0.5 M imidazole was added to each sample and HIS<sub>6</sub>-tagged proteins were eluted off the resin for 60 min at room temperature, with gentle agitation. Resin was spun down at 1,000 RCF for 1 min, and the supernatant was removed. 10  $\mu\text{l}$  of each sample was resolved on an SDS-PAGE gel alongside an input control and analysed by western blotting using the previously described antibody against MCM7<sup>49</sup>. In Extended Data Fig. 8d, a longer exposure of the IP lanes is shown, since they are far less intense than the input lanes.

45. Prokhorova, T. A., Mowrer, K., Gilbert, C. H. & Walter, J. C. DNA replication of mitotic chromatin in *Xenopus* egg extracts. *Proc. Natl Acad. Sci. USA* **100**, 13241–13246 (2003).
46. Lebofsky, R., Takahashi, T. & Walter, J. C. DNA replication in nucleus-free *Xenopus* egg extracts. *Methods Mol. Biol.* **521**, 229–252 (2009).
47. Budzowska, M., Graham, T. G. W., Sobek, A., Waga, S. & Walter, J. C. Regulation of the Rev1-pol  $\zeta$  complex during bypass of a DNA interstrand cross-link. *EMBO J.* **34**, 1971–1985 (2015).
48. Magnaghi, P. *et al.* Covalent and allosteric inhibitors of the ATPase VCP/p97 induce cancer cell death. *Nature Chem. Biol.* **9**, 548–556 (2013).
49. Walter, J. C. & Newport, J. Initiation of eukaryotic DNA replication: origin unwinding and sequential chromatin association of Cdc45, RPA, and DNA polymerase  $\alpha$ . *Mol. Cell* **5**, 617–627 (2000).
50. Long, D. T., Räschele, M., Joukov, V. & Walter, J. C. Mechanism of RAD51-dependent DNA interstrand cross-link repair. *Science* **333**, 84–87 (2011).

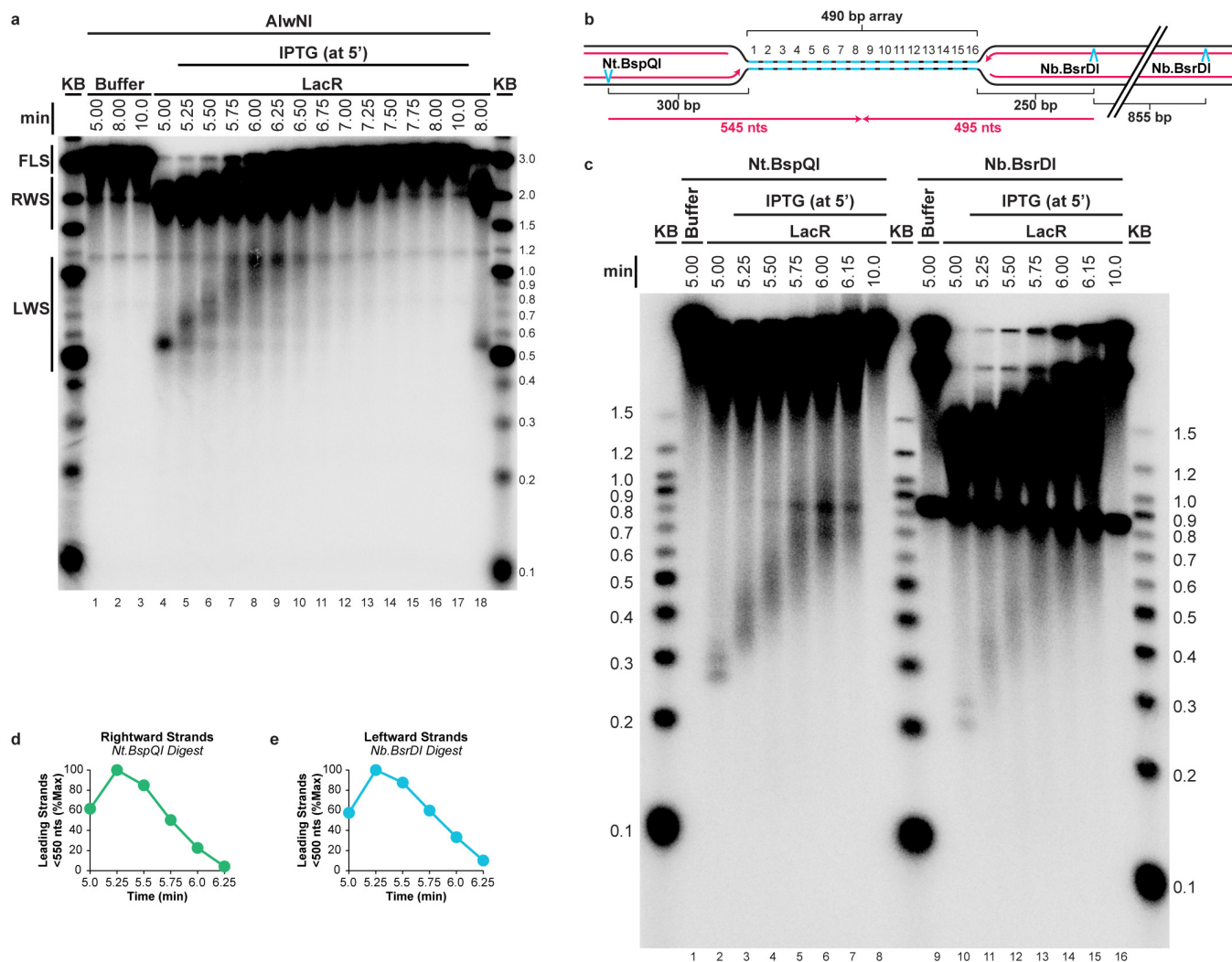




### Extended Data Figure 1 | Sequence-specific termination can be induced at a LacR array.

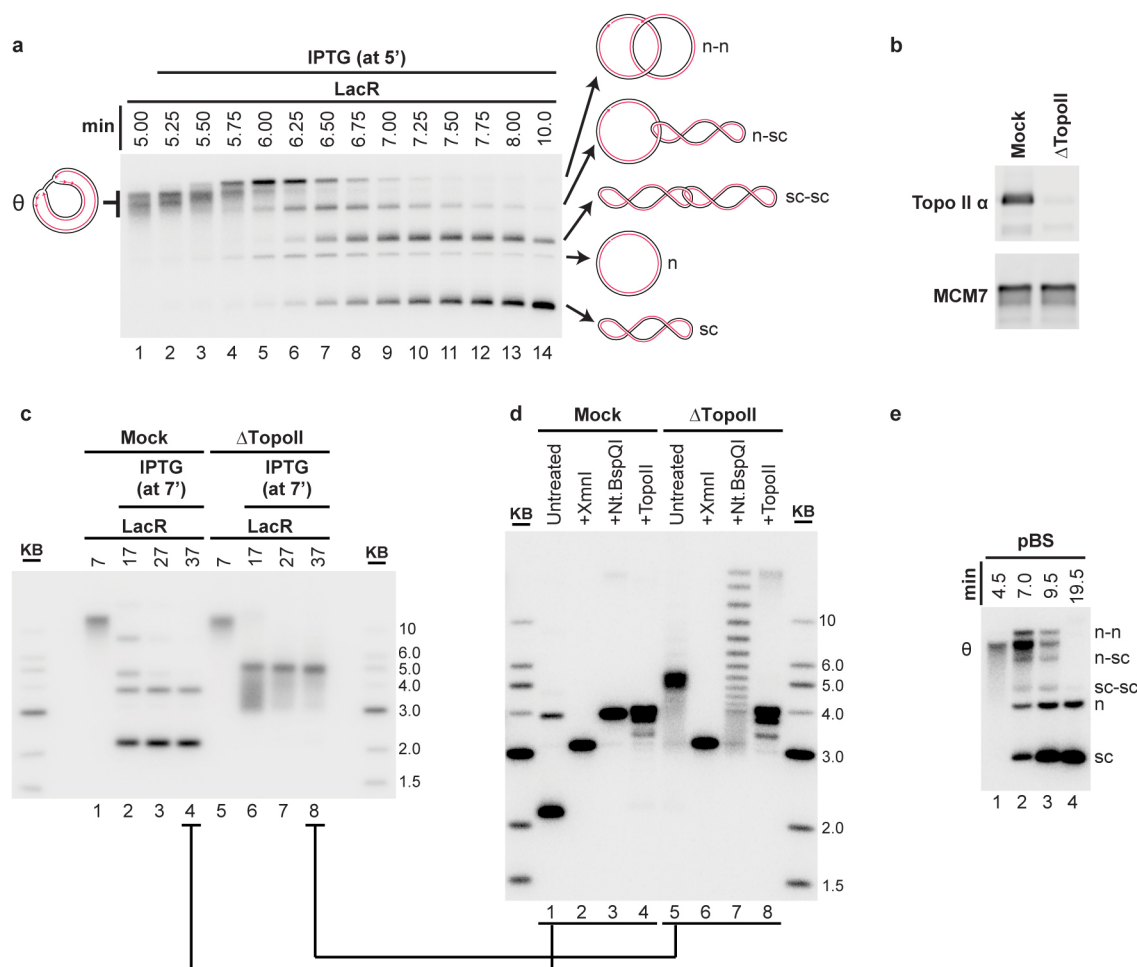
**a**, To investigate whether a LacR array blocks replication forks, a plasmid containing a tandem array of 16 *lacO* sequences, p[lacO<sub>16</sub>] (or p[lacOx16]), was incubated with buffer or LacR and then replicated in egg extract containing [ $\alpha$ -<sup>32</sup>P]dATP. Radiolabelled replication intermediates were cleaved with XmnI (far left cartoon) and separated according to size and shape by 2D gel electrophoresis (see schematic of 2D gel). As replication neared completion at 4.5 min, mainly linear molecules were produced in the presence of buffer (orange arrowhead). In contrast, in the presence of LacR, a discrete spot appeared on the double-Y arc (blue arrowhead), demonstrating that converging replication forks accumulate at a specific locus on p[lacO<sub>16</sub>]. These data indicate that 16 copies of LacR block replication forks. **b–f**, To test whether the double-Y structures observed in panel **a** arose from replication forks stalling at the outer edges of the *lacO* array, we tested whether LacR specifically inhibited replication of *lacO* sequences. To this end, p[lacO<sub>16</sub>] (**c**) and the parental plasmid lacking *lacO* repeats, p[empty] (**b**),

were incubated in the presence of buffer or LacR and replicated using *Xenopus* egg extracts containing [ $\alpha$ -<sup>32</sup>P]dATP. Radiolabelled replication intermediates were cleaved with AflIII and PvuII to release the 2,354-bp plasmid backbone (**b** and **c**) and a 294-bp control fragment from p[empty] (**b**) or a 794-bp *lacO* fragment from p[lacO<sub>16</sub>] (**c**). The plasmid backbone and the respective inserts were separated on a native gel and detected by autoradiography (**d**). A longer exposure of the small fragments is shown, since they are less intense than the large fragments. The results in panel **d** were quantified in **e** and **f**. Notably, LacR specifically inhibited replication of the *lacO*-containing fragment in p[lacO<sub>16</sub>] (**f**, blue circles) but not the control fragment in p[empty] (**e**, green circles). We conclude that LacR prevents replication of the *lacO* array and that the double-Ys in panel **a** represent forks converged on the outer edges of the array. Importantly, synthesis within the 2,354-bp backbone fragment (**f**, orange circles) of p[lacO<sub>16</sub>] was not inhibited in the presence of LacR, indicating that no global structural changes occur that inhibit replication.



**Extended Data Figure 2 | Supplementary fork progression data.** **a**, The gel shown in Fig. 1e was overexposed and shown in its entirety so that the smaller leftward strands (LWS, Fig. 1d) could be detected. As observed for the rightward strands (RWS, Fig. 1e), LWS rapidly increased in size and then disappeared as they were ligated to produce full-length strands (FLS, Fig. 1e). **b–e**, To determine whether the heterogeneity of LWS (**a**) and RWS (Fig. 1e) was due to delayed extension of lagging strands, or because a significant fraction of leading strands did not restart upon IPTG addition, we specifically monitored leading strand progression upon IPTG addition on *p[lacO<sub>16</sub>]*. To this end, DNA samples were treated with Nt.BspQI or Nb.BsrDI to specifically liberate the rightward or leftward leading strands, respectively (**b**), and DNAs were separated on a denaturing agarose gel (**c**). Before IPTG addition,

discrete leading strand products of the expected size were observed (lanes 2 and 10). The presence of two stall products reflects the fact that at a slow rate, the replisome bypasses LacR (see also Fig. 3). Upon IPTG addition, these species rapidly and completely shifted up the gel, indicating that rightward and leftward leading strands restarted efficiently. Therefore, the heterogeneity of the LWS (**a**) and RWS (Fig. 1e) is probably due to delayed ligation of a new Okazaki fragment to the lagging strands. Quantification of leading strands that had not reached the midpoint of the array (rightward and leftward strands smaller than 550 and 500 nt, respectively, **b**) revealed that by 6.25 min, 90% of rightward and leftward leading strands passed the midpoint of the array (**d, e**). This demonstrates that leading strands pass each other when forks meet. KB, kilobase ladder, with the length of each band (in kilobases) labelled.



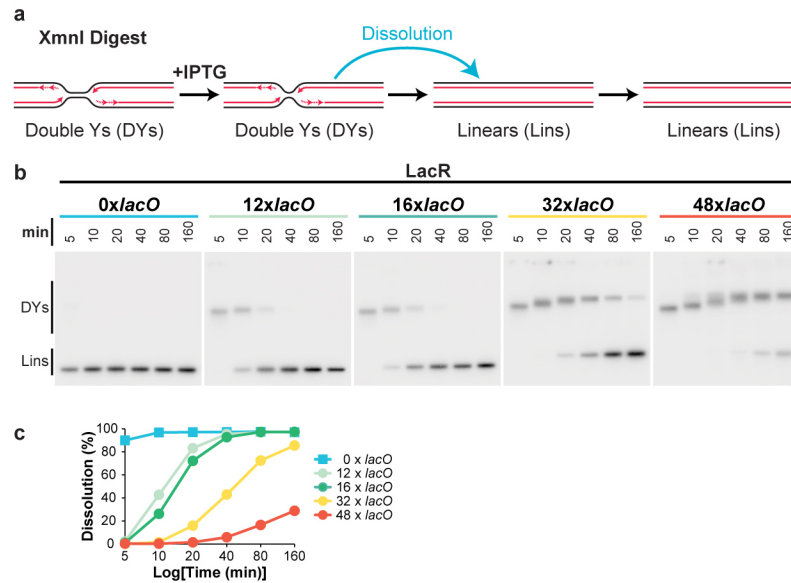
### Extended Data Figure 3 | Topo-II-dependent decatenation of p[lacO<sub>16</sub>].

**a**, The autoradiograph in primary Fig. 1g is reproduced with cartoons indicating the structures of the replication and termination intermediates n–n, n–sc, sc–sc, n and sc (see Fig. 1 for definitions). The order of appearance of the different catenanes matches previous work<sup>5</sup> (n–n, then n–sc, then sc–sc).

**b–d**, To determine the role of Topo II during termination within a *lacO* array, termination was monitored in mock- or Topo-II-depleted extracts. To confirm immunodepletion of Topo II, mock and Topo-II-depleted NPE was blotted with MCM7 and Topo II antibodies (**b**). p[lacO<sub>16</sub>] was incubated with LacR, then replicated in either mock- or Topo-II-depleted egg extracts in the presence of [ $\alpha$ -<sup>32</sup>P]dATP, and termination was induced with IPTG (at 7 min). Untreated DNA intermediates were separated by native gel electrophoresis (**c**). In the mock-depleted extract, nicked and supercoiled monomers were readily produced (as in panel **a**, albeit with slower kinetics due to nonspecific inhibition of the extracts by the immunodepletion procedure), while in the Topo-II-depleted extracts, a discrete species was produced. DNA from the last time point in each reaction (lanes 4 and 8 in panel **c**) was purified and treated with XmnI, which cuts p[lacO<sub>16</sub>] once, or Nt.BspQI, which nicks p[lacOx16] once, or recombinant Topo II, and then separated by native gel electrophoresis (**d**). Cleavage of the mock- and Topo-II-depleted products with XmnI yielded the expected linear 3.15-kb band (lanes 2 and 6), demonstrating that

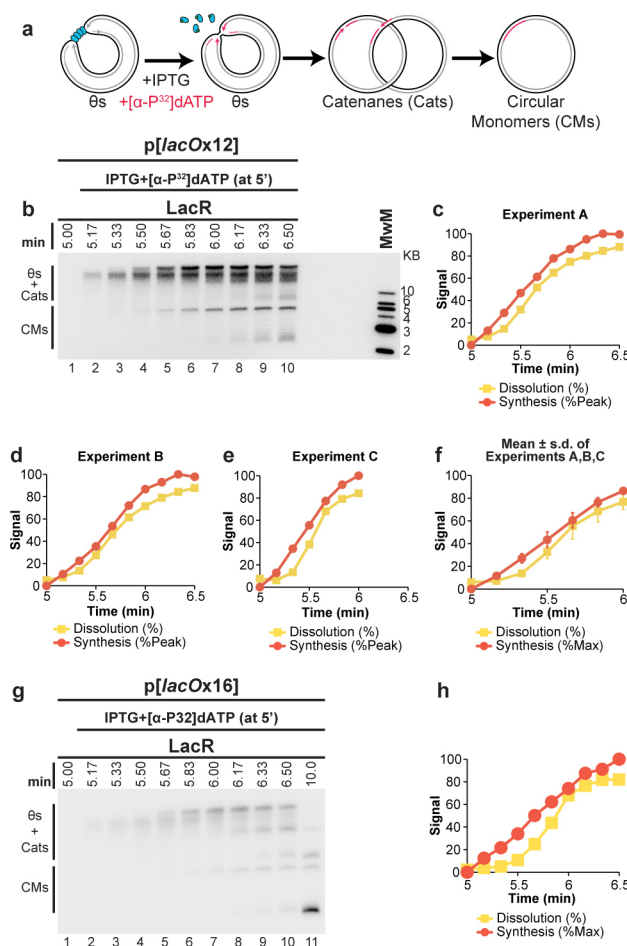
in both extracts all products were fully dissolved topoisomers of each other. Relaxation of the mock-depleted products by nicking with Nt.BspQI yielded a discrete band corresponding to nicked plasmid (lane 3), while the Topo-II-depleted products were converted to a ladder of discrete topoisomers (lane 7), which we infer represent catenated dimers of different linking numbers, since the mobility difference cannot be due to differences in supercoiling. Importantly, the mobility shift after Nt.BspQI treatment (lane 5 versus lane 7) demonstrated that the Topo-II-depleted products (lane 5) were covalently closed and thus in the absence of Topo II, ligation of the daughter strands still occurred. Treatment of the mock- and Topo-II-depleted products with recombinant human Topo II produced the same relaxed monomeric species (lanes 4 and 8), further confirming that the Topo-II-depleted products contained catenanes. Collectively, these observations demonstrate that termination within a *lacO* array in Topo-II-depleted extracts produces highly catenated supercoiled–supercoiled dimers, as seen in cells lacking Topo II<sup>16,17</sup>. These data confirm that Topo II is responsible for decatenation and argue that termination within a *lacO* array reflects physiological termination. **e**, n–n, n–sc, sc–sc, n and sc products were also detected when plasmid lacking *lacO* sequences (pBlueScript) was replicated in the absence of LacR without the use of cyclin A to synchronize replication. Therefore, these intermediates arise in the course of unperturbed DNA replication in *Xenopus* egg extracts.





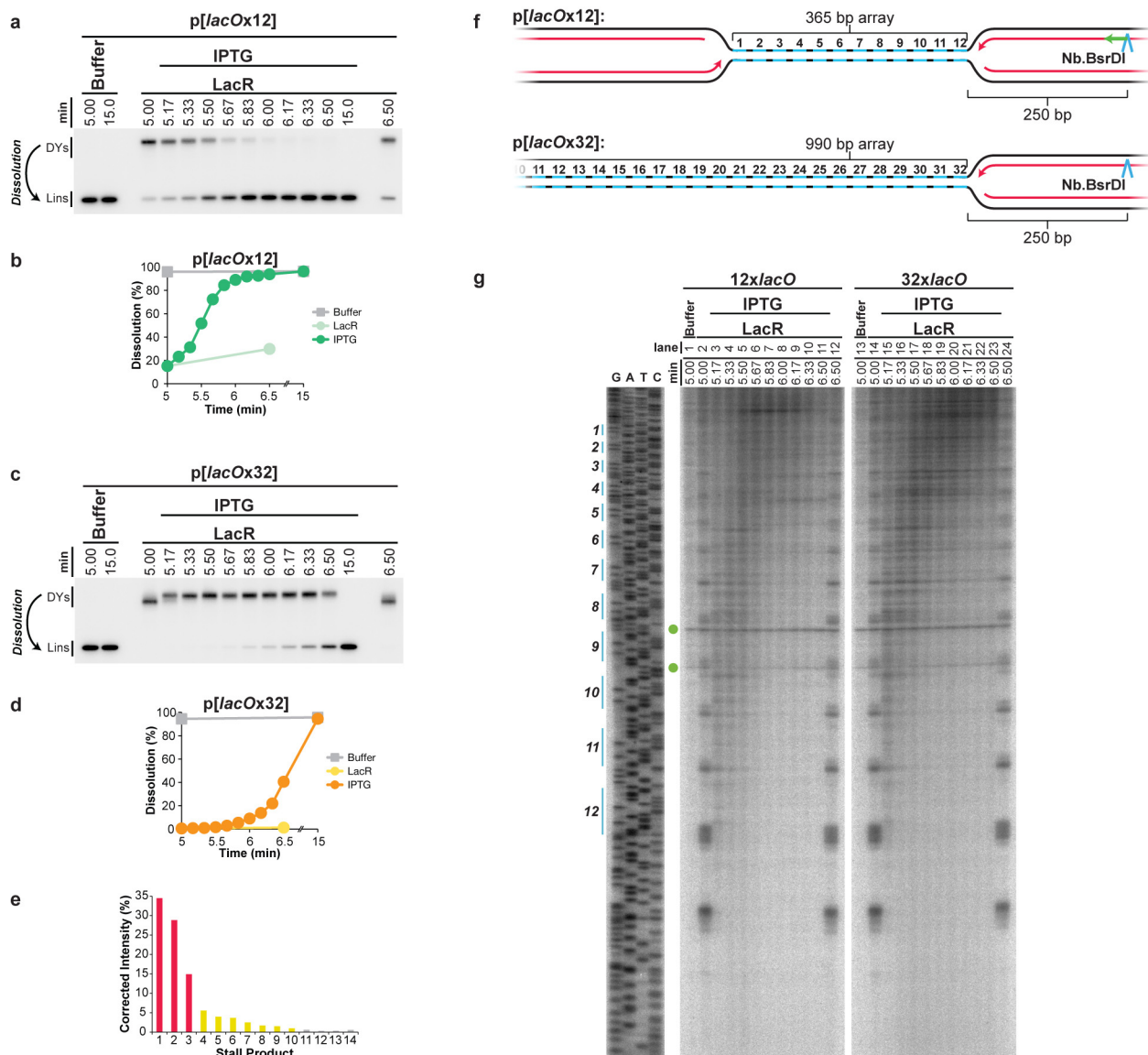
**Extended Data Figure 4 | Inhibition of termination by different-sized LacR arrays.** **a**, Cartoon depicting intermediates detected in the dissolution assay. **b**, To determine the ability of different-sized LacR arrays to inhibit termination, the earliest stage of termination, dissolution (**a**), was monitored in plasmids containing 0, 12, 16, 32, or 48 *lacO* repeats. Plasmids were incubated with LacR, and replicated in the presence of [ $\alpha$ - $^{32}$ P]dATP. To measure dissolution, radiolabelled termination intermediates were cut with XmnI. Cleaved products

were separated on a native agarose gel and detected by autoradiography. **c**, Quantification of dissolution in **b**. When 12 or more *lacO* repeats were present in the array, dissolution was robustly inhibited for at least 5 min. Potent inhibition lasted 10 min when 32 *lacO* sequences were present, and 20 min in the presence of 48 *lacO* sequences. In the absence of *lacO* sequences, dissolution was essentially complete by 5 min. Therefore, 12 *lacO* repeats are sufficient to inhibit termination for 5 min.



**Extended Data Figure 5 | The rate of total DNA synthesis does not slow before dissolution.** **a–c**, To test further whether replication stalls or slows before dissolution, p[lacO<sub>12</sub>] was pre-incubated with LacR and replicated in *Xenopus* egg extracts. Termination was then induced by addition of IPTG after 5 min. Simultaneously, [ $\alpha$ -<sup>32</sup>P]dATP was added to specifically radiolabel DNA synthesized after IPTG addition (**a**). Radiolabelled DNA was then separated on a native agarose gel and total signal was measured by autoradiography (**b**). Total signal was quantified, normalized to peak signal, and graphed alongside the rate of dissolution, which was also measured in the same experiment (**c**). This approach gives a highly sensitive measure of DNA synthesis without manipulation of DNA samples. DNA synthesis should occur primarily within the *lacO* array (see Extended Data Fig. 1). Upon IPTG addition, there was an approximately linear increase in signal, which plateaued by 5.83 min. Importantly, dissolution was 65% complete by 5.83 min. Therefore, the large majority of dissolution occurs without stalling of DNA

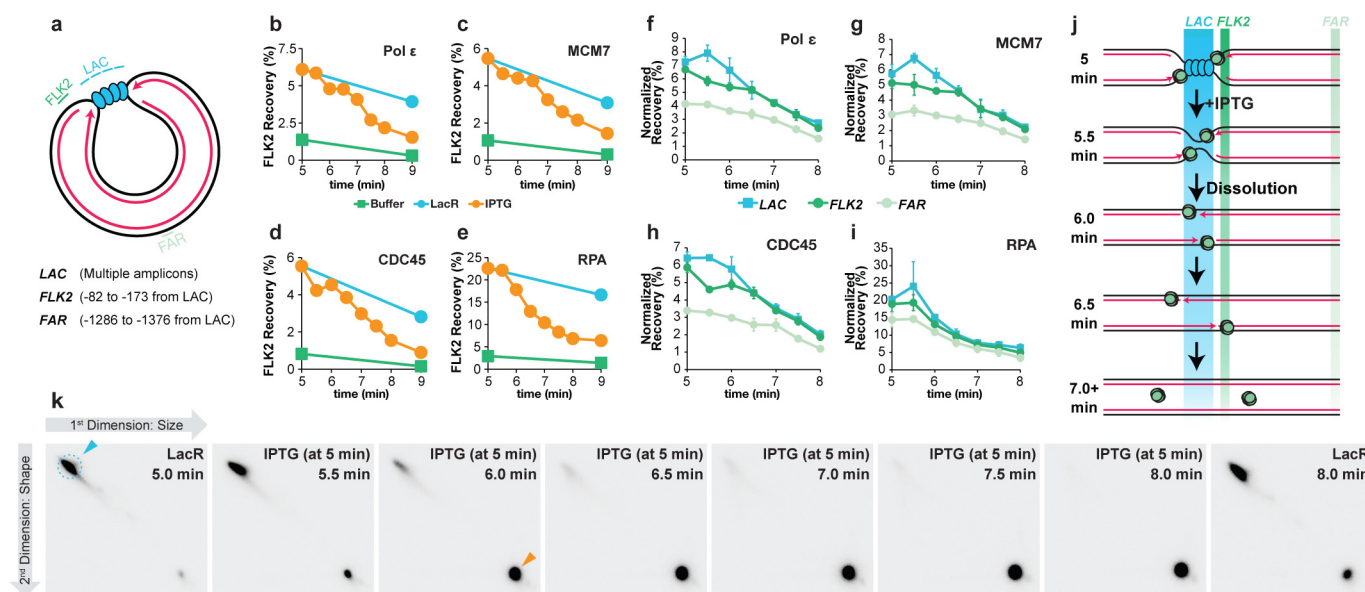
synthesis. **d, e**, Experimental repeats of **b, c**. **f**, The experiments shown in **c–e** were graphed together with mean  $\pm$  s.d. Synthesis data were normalized so that for each experiment, synthesis at 1 min was assigned a value of 84.4%, since this was the average value from **c, d**, where synthesis was allowed to plateau. Given the rate of replication fork progression in these egg extracts (260 bp min<sup>-1</sup> (ref. 32)) and the size of the array (365 bp), forks should require, on average, 0.7 min to converge if no stalling occurs ((365 bp/2)/260 bp min<sup>-1</sup> = 0.7 min). The time required for dissolution was not appreciably longer than this (dissolution was 50% complete by 0.67 min after IPTG addition, **f**), consistent with a lack of stalling. **g, h**, The experiment shown in **b, c** was repeated using p[lacO<sub>16</sub>]. Synthesis was approximately linear until 6.17 min, at which point 81% of molecules had dissolved, further demonstrating that the majority of dissolution occurs without stalling of DNA synthesis.



**Extended Data Figure 6 | Replisome progression through 12 and 32 *lacO* arrays.** **a–d**, To test whether replisomes meet later in a *lacO*<sub>32</sub> array than a *lacO*<sub>12</sub> array, we monitored dissolution. LacR block-IPTG release was performed on *p[lacOx12]* and *p[lacOx32]* and radiolabelled termination intermediates were digested with XmnI to monitor the conversion of double-Y molecules to linear molecules (dissolution). Cleaved molecules were separated on a native agarose gel, detected by autoradiography (**a**, **c**), and quantified (**b**, **d**). Upon IPTG addition, dissolution was delayed by at least 1 min within the 32 *lacO* array compared to the 12 *lacO* array (**b**, **d**). Moreover, by 6 min, 92% of forks had undergone dissolution on *p[lacOx12]* while only 9% had dissolved on *p[lacOx32]* (**b**, **d**). **e**, Stall products within the 12 *lacO* array (Fig. 3b, lane 2) were quantified, signal was corrected based on size differences of the products, and the percentage of stall products at each stall point was calculated. 78% of leading strands stalled at the first three arrest points (red columns), 19% stalled at the fourth to tenth arrest points (yellow columns) and the remaining 3% stalled at the tenth to fourteenth arrest points (grey columns). The appearance of fourteen arrest points is reproducible but surprising, given that the presence of only 12 *lacO* sequences was confirmed by sequencing in the very preparation of *p[lacOx12]* that was used in Fig. 3. The thirteenth and

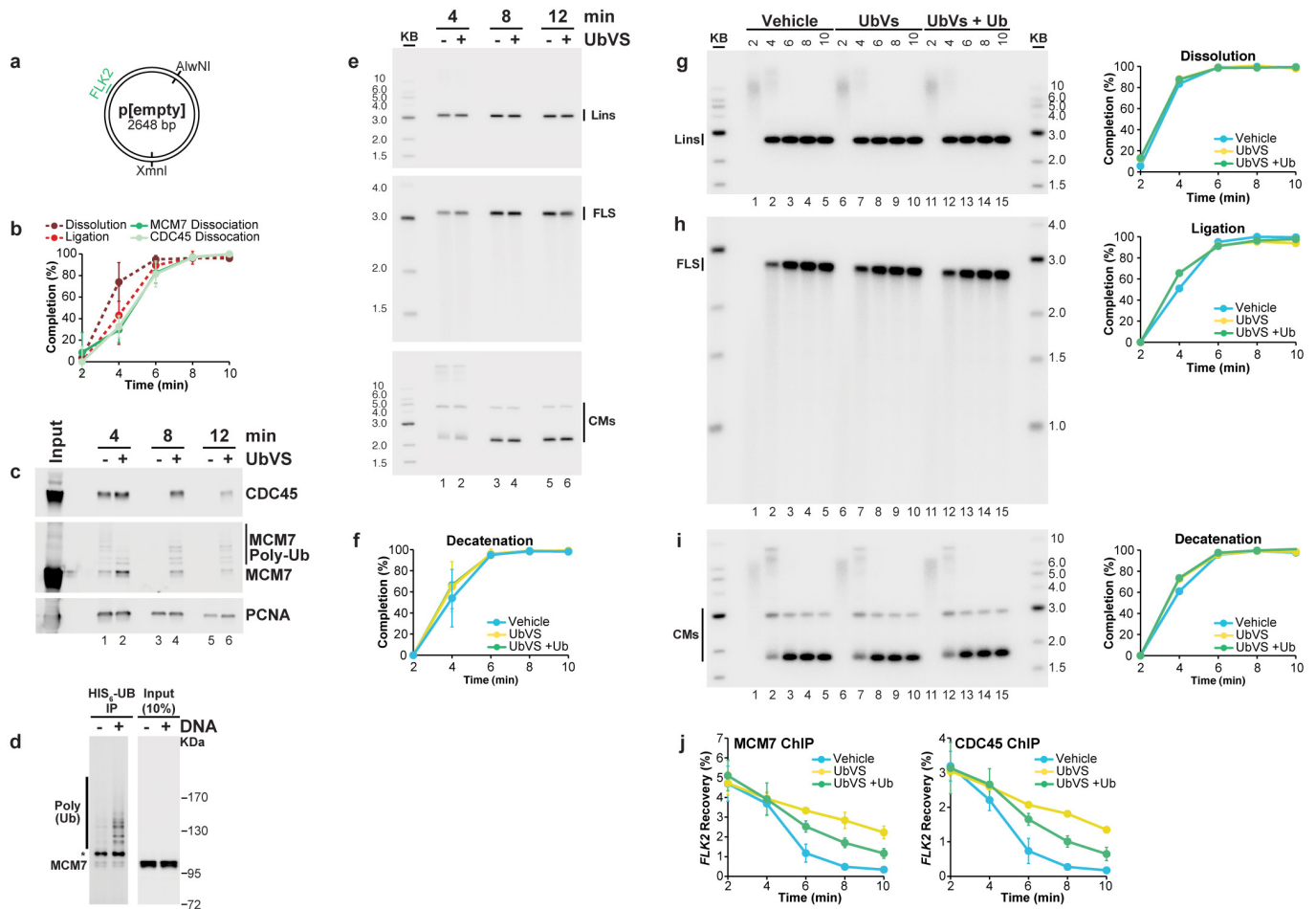
fourteenth arrest points cannot stem from cryptic *lacO* sites beyond the twelfth *lacO* site, as this would position the first leftward leading strand stall product ~90 nucleotides from the *lacO* array, instead of the observed ~30 nucleotides (see **f**, **g**). At present, we do not understand the origin of these stall products. **f**, **g**, Progression of leftward leading strands into the array. The same DNA samples used in Fig. 3 were digested with the nicking enzyme Nb.BsrDI, which released leftward leading strands (**f**), and separated on a denaturing polyacrylamide gel (**g**). The *lacO* sites of *p[lacOx12]* are highlighted in blue on the sequencing ladder (**g**), which was generated using the primer JDO109 (green arrow, **f**). Green circles indicate two nonspecific products of digestion. These products arise because nicking enzyme activity varies between experiments, even under the same conditions. There was no significant difference in the pattern of leftward leading strand progression between the 12 *lacO* and 32 *lacO* arrays, as seen for the rightward leading strands (Fig. 3b). Specifically, by 5.67 min, the majority of leading strands had extended beyond the seventh *lacO* repeat within *lacO*<sub>12</sub> (lane 6) and the equivalent region of *lacO*<sub>32</sub> (lane 18). Therefore, progression of leftward leading strands is unaffected by the presence of an opposing replisome, suggesting that converging replisomes do not stall when they meet.





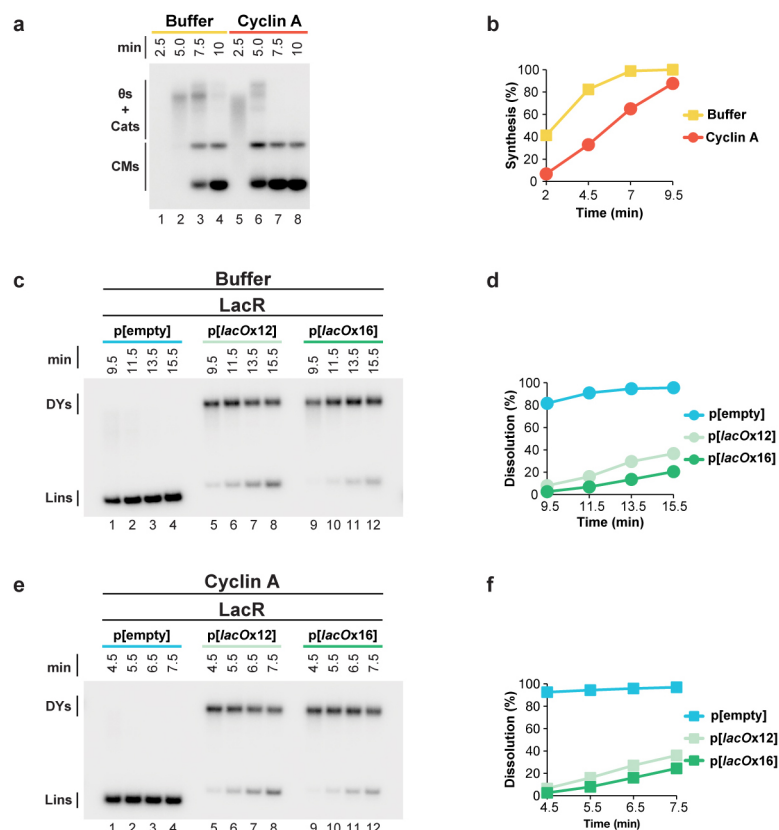
**Extended Data Figure 7 | Supplementary ChIP data.** **a**, Cartoon depicting the LAC, FLK2 and FAR loci, which were used for ChIP. Their precise locations relative to the leftward edge of the *lacO* array are indicated. The LAC amplicon is present in four copies distributed across the *lacO*<sub>16</sub> array and three copies distributed across the *lacO*<sub>12</sub> array. **b–e**, p[*lacOx12*] was incubated with buffer or LacR and termination was induced at 5 min by IPTG addition. MCM7, RPA, CDC45 and Pole ChIP was performed at different time points after IPTG addition but also in the buffer control and no IPTG control. Recovery of FLK2 was measured as a percentage of input DNA. Upon IPTG addition, ChIP signal declined and by 9 min was comparable to the buffer control, demonstrating that unloading of replisomes was induced within 4 min of IPTG addition. **f**, To test whether movement of the replisome into and out of the *lacO* array could be detected upon IPTG addition, termination was monitored within a *lacO* array, and we performed ChIP of the leading strand polymerase Pole, which was inferred to move into and out of the array based on the behaviour of leading strands during termination (Extended Data Fig. 2b–e). It was predicted that Pole ChIP at the LAC locus should increase slightly as Pole enters the *lacO* array and decline again as converging polymerases pass each other, but persist at FLK2 while the polymerases move out of the array. Before IPTG addition, Pole was enriched at LAC and FLK2 compared to FAR, consistent with the leading strands being positioned on either side of the *lacO* array (Extended Data Fig. 2c and Fig. 3). Upon IPTG addition, Pole became modestly enriched at LAC compared to FLK2 (5.5 min) but then declined to similar levels at both LAC and FLK2 by 6.5 min. These data are consistent with the leading strand polymerases entering the *lacO* array and passing each other. **g, h**, To test whether CMG exhibited the same ChIP profile as Pole, MCM7 and CDC45 ChIP was performed using the same samples. After IPTG addition, MCM7 and CDC45 were enriched at LAC compared to FLK2 (5.5 min), then declined to similar levels at both LAC and FLK2 by 6.5 min, as seen for Pole (f). These data are consistent with a model in which CMGs enter the array and pass each other during termination. A caveat of these experiments is the relatively high recovery of the FAR locus in MCM7, CDC45 and Pole ChIP. Specifically, signal was at most only ~2-fold enriched at LAC compared to FAR. This was not due to high background binding, because by the end of the experiment (10 min time point, not shown) we observed a decrease in signal of ~5–7-fold. Furthermore, we observed ~5–7-fold enrichment in binding (ChIP) of replisome components to p[*lacO*<sub>12</sub>] that had been incubated in LacR compared

to a buffer control (see g–i, below). Instead, the high FAR signal was probably due to poor spatial resolution of the ChIP. Consistent with this, when a plasmid containing a DNA interstrand cross-link (ICL) was replicated, essentially all replisomes converged upon the ICL but the ChIP signal for MCM7 and CDC45 was only ~3–4-fold enriched at the ICL compared to a control locus<sup>41</sup>. We speculate that the higher background observed at the control locus in our experiments is due to the decreased distance of the control locus from the experimental locus (1.3 kb for p[*lacO*<sub>16</sub>] and p[*lacO*<sub>12</sub>] versus 2.4 kb for the ICL plasmid) and possibly due to increased catenation of the parental strands during termination. The high signal at FAR should not complicate interpretation of the MCM7, CDC45 and Pole ChIP (f), as signal at FAR was essentially unaltered between 5 and 6.5 min. Further evidence that the high signal seen at the FAR locus emanates from forks stalled near the *lacO* array is presented in panel k. **i**, ChIP of RPA was performed on the same chromatin samples used in b–d. As seen for Pole, MCM7 and CDC45, enrichment of RPA at LAC compared to FAR was relatively low, consistent with poor spatial resolution. **j**, Predicted binding of CMGs to the LAC, FLK2 and FAR loci before and after IPTG addition if converging CMG pass each other. **k**, To determine whether most forks stalled at the array and not elsewhere in the plasmid, we performed a time course in which p[*lacO*<sub>16</sub>] undergoing termination was examined by 2D gel electrophoresis at various time points. p[*lacO*<sub>16</sub>] was pre-bound to LacR and replicated in *Xenopus* egg extract containing [ $\alpha$ -<sup>32</sup>P]dATP. Termination was induced by IPTG addition and samples were withdrawn at different times. Radiolabelled replication intermediates were cleaved with XmnI (as in Extended Data Fig. 1a) and separated according to size and shape on 2D gels<sup>50</sup>. A parallel reaction was performed in which samples were analysed by ChIP, which was one of the repeats analysed in b–e. In the presence of LacR, a subset of double-Y molecules accumulated (blue arrowhead), demonstrating that 83% of replication intermediates (signal in dashed blue circle) contained two forks converged at a specific locus. After IPTG addition, linear molecules rapidly accumulated (orange arrowhead) as dissolution occurred. Importantly, the vast majority of signal was present in the discrete double-Y and linear species (blue and orange arrowheads), demonstrating that the relatively high ChIP signal observed at FAR in panels f–i was derived from forks present at the *lacO*<sub>16</sub> array and not elsewhere.



**Extended Data Figure 8 | Supplementary termination data for p[empty] experiments.** **a**, Cartoon depicting the *XmnI* and *AlwNI* sites on p[empty], which are used for the dissolution and ligation assays, respectively, and the *FLK2* locus, which is used for ChIP. **b**, Plasmid DNA without a *lacO* array (p[empty]) was replicated and at different times chromatin was subjected to MCM7 and CDC45 ChIP. Per cent recovery of *FLK2* was quantified and used to measure dissociation of MCM7 and CDC45 (see Methods). Dissolution and ligation were also quantified in parallel. Mean  $\pm$  s.d. is plotted ( $n = 3$ ). The MCM7 and CDC45 dissociation data are obtained from the vehicle controls in Fig. 5b, c, while the dissolution and ligation data are obtained from the vehicle controls in Fig. 5d, e. **c**, To seek independent evidence for the conclusions of the ChIP data presented in Fig. 5b, c, we used a plasmid pull-down procedure. p[empty] was replicated in egg extracts treated with vehicle or Ub-VS. At the indicated times, chromatin-associated proteins were captured on LacR-coated beads (which binds DNA independently of *lacO* sites) and analysed by western blotting for CDC45, MCM7 and PCNA. CDC45 and MCM7 dissociated from chromatin by 8 min in the vehicle control, but persisted following Ub-VS treatment. **d**, To test whether the MCM7 modifications detected in panel c represented ubiquitylation, extracts were incubated with His<sub>6</sub>-ubiquitin in the absence of cyclin A, and in the absence or presence of plasmid DNA. After 15 min, His<sub>6</sub>-tagged proteins were captured

by nickel resin pull down and blotted for MCM7. DNA replication greatly increased the levels of ubiquitylated MCM7, with the exception of a single species that was ubiquitylated independently of DNA replication (\*). These data show that MCM7 is ubiquitylated during plasmid replication in egg extracts, as observed in yeast and during replication of sperm chromatin after nuclear assembly in egg extracts<sup>24,25</sup>. **e**, In parallel to the plasmid pull downs performed in **c**, DNA samples were withdrawn for dissolution, ligation and decatenation assays, none of which was perturbed by Ub-VS treatment. These data support our conclusion, based on ChIP experiments (Fig. 5), that defective CMG unloading does not affect dissolution, ligation, or decatenation. **f**, Decatenation was measured in the same reactions used to measure dissolution and ligation (Fig. 5d, e), mean  $\pm$  s.d. is plotted ( $n = 3$ ). **g–i**, Given the experimental variability at the 4 min time point in Fig. 5d–f, the primary data and quantification for dissolution (**g**), ligation (**h**) and decatenation (**i**) for one of the three experiments summarized in Fig. 5d–f is presented. This reveals that Ub-VS does not inhibit dissolution, ligation, or decatenation at the 4 min time point. The same conclusion applies to two additional repetitions of this experiment (data not shown). **j**, The primary ChIP data used to measure dissociation of MCM7 and CDC45 in Fig. 5b, c is shown. Recovery of *FLK2* was measured. Mean  $\pm$  s.d. is plotted ( $n = 3$ ).



**Extended Data Figure 9 | Cyclin A treatment synchronizes DNA replication in *Xenopus* egg extracts.** **a, b**, To synchronize DNA replication in *Xenopus* egg extracts, we treated extracts with cyclin A, which probably accelerates replication initiation<sup>45</sup>. Plasmid DNA was incubated in High Speed Supernatant for 20 min, then either buffer or cyclin A was added for a further 20 min. NucleoPlasmic extract was added to initiate DNA replication, along with [ $\alpha$ -<sup>32</sup>P]dATP to label replication intermediates. Replication products were separated on a native agarose gel, detected by autoradiography (**a**), and quantified (**b**). In the presence of vehicle, replication was not complete by 9.5 min, but in the presence of cyclin A, replication was almost complete by 4.5 min (**b**). Thus, cyclin A treatment approximately doubles the speed of DNA replication in *Xenopus* egg extracts. **c–f**, To test whether cyclin A affects the ability of LacR to inhibit termination, we monitored dissolution of plasmids containing a 12 or 16 LacR array in the presence and absence of cyclin A. p[lacO<sub>12</sub>], p[lacO<sub>16</sub>], and the parental control plasmid p[empty] were incubated with LacR, and then treated with buffer or cyclin A before replication was initiated with NPE in the presence of [ $\alpha$ -<sup>32</sup>P]dATP. Samples were withdrawn

when dissolution of p[empty] plateaued (9.5 min in the presence of buffer, 4.5 min in the presence of cyclin A). Given that cyclin A treatment approximately doubles the speed of replication (see **b**), samples were withdrawn from these reactions twice as frequently as the buffer-treated samples. To measure dissolution, radiolabelled termination intermediates were cut with XmnI to monitor the conversion of double-Y molecules to linear molecules. Cut molecules were separated on a native agarose gel and detected by autoradiography (**c, e**). By the time the first sample was withdrawn, dissolution of p[empty] was essentially complete, in the absence (9.5 min, **d**) or presence (4.5 min, **f**) of cyclin A. Importantly, dissolution of p[lacO<sub>12</sub>] and p[lacO<sub>16</sub>] was prevented in the absence (9.5 min, **d**) or presence (4.5 min, **f**) of cyclin A. Moreover, dissolution occurred approximately twice as fast in the presence of cyclin A (note the similarity between **d** and **f** even though samples are withdrawn twice as frequently in **f**) consistent with replication being approximately twice as fast in the presence of cyclin A. Therefore, cyclin A does not affect the ability of a LacR array to block replication forks.



Extended Data Table 1 | Tables of plasmids and oligonucleotides used

**A.**

Plasmid	Insert	Construction
pJD82	SacI-BsrGI-(lacO)x4-BsiWI-KpnI	Replacement of the sequence between SacI and KpnI of pBluescript II KS-
pJD85	SacI-BsrGI-(lacO)x8-BsiWI-KpnI	JDO38/39 annealed and cloned into pJD82 that had been cut with BsrGI
pJD88	SacI-BsrGI-(lacO)x16-BsiWI-KpnI	BsrGI/BsiWI fragment from pJD85 cloned into pJD85 that had been cut with BsrGI
pJD92	SacI-BsrGI-(lacO)x32-BsiWI-KpnI	BsrGI/BsiWI fragment from pJD88 cloned into pJD88 that had been cut with BsrGI
pJD100	SacI-BsrGI-(lacO)x48-BsiWI-KpnI	BsrGI/BsiWI fragment from pJD88 cloned into pJD92 that had been cut with BsrGI
pJD104	SacI-BsrGI-(lacO)x12-BsiWI-KpnI	JDO38/39 annealed and cloned into pJD85 that had been cut with BsrGI
pJD105	SacI-Nb.BsmI-BsiWI-Nt.BbvCI-KpnI	Replacement of the sequence between SacI and KpnI of pBluescript II KS- with JDO42/43
pJD139	SacI-Nb.BsmI-BsiWI-Nt.BbvCI-KpnI	Quickchange mutagenesis of pJD105 using JDO94/95
pJD145	SacI-Nb.BsmI-BsiWI-Nt.BbvCI-KpnI	Quickchange mutagenesis of pJD139 using JDO100/10
pJD150	SacI-Nb.BsmI-(lacO)x12-BsiWI-Nt.BbvCI-KpnI	BsrGI/BsiWI fragment from pJD104 cloned into pJD145 that had been cut with BsiWI
pJD152	SacI-Nb.BsmI-(lacO)x16-BsiWI-Nt.BbvCI-KpnI	BsrGI/BsiWI fragment from pJD88 cloned into pJD145 that had been cut with BsiWI
pJD156	SacI-Nb.BsmI-(lacO)x32-BsiWI-Nt.BbvCI-KpnI	BsrGI/BsiWI fragment from pJD92 cloned into pJD145 that had been cut with BsiWI
pQNT	-	pCDFDuet-1 containing a HincII site (pQuant from <sup>41</sup> )

**B.**

Oligo	Sequence	Description
JDO38	5'-GTACATCAATTGTGAGCGGATAACAATTGTGTA GGGAGGAATTGTGAGCGGATAACAATTGGAGTTG ATAATTGTGAGCGGATAACAATTGGCTTCAACGTA ATTGTGAGCGGATAACAATTTC-3'	Can be annealed to JDO39 to generate dsDNA containing 4x lacO sites with ends that are compatible with BsiWI and BsrGI.
JDO39	5'-GTACGGAAATTGTTATCCGCTCACAATTACGT TGAAGCCAATTGTTATCCGCTCACAATTATCAACT CCAAATTGTTATCCGCTCACAATTCCTCCCTAACA ATTGTTATCCGCTCACAATTGAT-3'	Can be annealed to JDO38 to generate dsDNA containing 4x lacO sites with ends that are compatible with BsiWI and BsrGI.
JDO42	5'-CTGTACAGCATTCCCATGGCGTACGTTCTAGA CCTCAGCTATGGTACC-3'	Can be annealed to JDO43 to generate dsDNA containing sites for BsrGI-Nb.BsmI-NcoI-BsiWI-XbaI-Nb.BbvCI with ends that are compatible with SacI and KpnI.
JDO43	5'-AGCGGTACCATAGCTGAGGTCTAGAACGTACG CCATGGGAATGCTGTACAGAGCT-3'	Can be annealed to JDO42 to generate dsDNA containing sites for BsrGI-Nb.BsmI-NcoI-BsiWI-XbaI-Nb.BbvCI with ends that are compatible with SacI and KpnI.
JDO94	5'-TAAGGGATTTTGCCGATTTCGGCCTATGCTCT TCGCACTGTGGTTAAAAATGAGC-3'	Used with JDO95 to introduce Nt.BspQI and Nb.BtsI sites upstream of Nb.BsmI in pJD105-derived plasmids by Quickchange mutagenesis.
JDO95	5'-GCTCATTTTTTAACCACTGCGAAGAGCATA GGCCGAAATCGGCAAAATCCCTTA-3'	Used with JDO94 to introduce Nt.BspQI and Nb.BtsI sites upstream of Nb.BsmI in pJD105-derived plasmids by Quickchange mutagenesis.
JDO100	5'-TGAGCGTCGATTCATTGCTTTGTGATGCTCGT CAGGGG-3'	Used with JDO101 to introduce Nb.BsrDI site downstream of BbvCI in pJD105-derived plasmids by Quickchange mutagenesis.
JDO101	5'-CCCCCTGACGAGCATCACAAGCAATGAATCG ACGCTCA-3'	Used with JDO100 to introduce Nb.BsrDI site downstream of BbvCI in pJD105-derived plasmids by Quickchange mutagenesis.
JDO107	5'-CAGTGTGGTTAAAAATGAGCTG-3'	Sequencing primer for mapping leading strands released by Nt.BspQI digestion
JDO109	5'-CATTGCTTTGTGATGCTCGT-3'	Sequencing primer for mapping leading strands released by Nb.BsrDI digestion
JDO110	5'-TGGTTAAAAATGAGCTGATTAAACA-3'	Sequencing primer for mapping lagging strands released Nb.BtsI digestion.
JDO111	5'-TGAGGTCTAGAACGTACGGA-3'	Sequencing primer for mapping leading strands released by Nb.BbvCI digestion.
FLK2_F	5'-TCTTCGCTATTACGCCAGCT-3'	Used with FLK2_R to amplify the region 82-173 bases upstream of the lacO array in pJD152
FLK2_R	5'-TTACAACGTCGTGACTGGGA-3'	Used with FLK2_F to amplify the region 82-173 bases upstream of the lacO array in pJD152
LAC_F	5'-AGCGGATAACAATTGTTAGGGA-3'	Used with LAC_R to amplify four sites within the lacO array in pJD152
LAC_R	5'-CTCACAATTACGTTGAAGCCAA-3'	Used with LAC_F to amplify four sites within the lacO array in pJD152
FAR_F	5'-ATTGCTACAGGCATCGTGGT-3'	Used with FAR_R to amplify the region 1286-1375 bases upstream of the lacO array in pJD152
FAR_R	5'-GGGATCATGTAACGCGCTTGA-3'	Used with FAR_F to amplify the region 1286-1375 bases upstream of the lacO array in pJD152
QNT_F	5'-TACAAATGTACGGCCAGCA-3'	Used with QNT_R to amplify pQNT
QNT_R	5'-GAGTATGAGGGAAGCGGTGA-3'	Used with QNT_F to amplify pQNT

# Relativistic boost as the cause of periodicity in a massive black-hole binary candidate

Daniel J. D'Orazio<sup>1</sup>, Zoltán Haiman<sup>1</sup> & David Schiminovich<sup>1</sup>

Because most large galaxies contain a central black hole, and galaxies often merge<sup>1</sup>, black-hole binaries are expected to be common in galactic nuclei<sup>2</sup>. Although they cannot be imaged, periodicities in the light curves of quasars have been interpreted as evidence for binaries<sup>3–5</sup>, most recently in PG 1302-102, which has a short rest-frame optical period of four years (ref. 6). If the orbital period of the black-hole binary matches this value, then for the range of estimated black-hole masses, the components would be separated by 0.007–0.017 parsecs, implying relativistic orbital speeds. There has been much debate over whether black-hole orbits could be smaller than one parsec (ref. 7). Here we report that the amplitude and the sinusoid-like shape of the variability of the light curve of PG 1302-102 can be fitted by relativistic Doppler boosting of emission from a compact, steadily accreting, unequal-mass binary. We predict that brightness variations in the ultraviolet light curve track those in the optical, but with a two to three times larger amplitude. This prediction is relatively insensitive to the details of the emission process, and is consistent with archival ultraviolet data. Follow-up ultraviolet and optical observations in the next few years can further test this prediction and confirm the existence of a binary black hole in the relativistic regime.

Assuming PG 1302-102 is a binary, it is natural to attribute its optical emission to gas that is bound to each black hole, forming circumpriary and circumssecondary accretion flows. Such flows, which form ‘minidisks’, are generically found in high-resolution two- and three-dimensional hydrodynamic simulations that include the black holes in their simulated domain<sup>8–15</sup>. Assuming a circular orbit, the velocity of the lower-mass secondary black hole is

$$v_2 = \left( \frac{2\pi}{1+q} \right) \left( \frac{GM}{4\pi^2 P} \right)^{1/3} = 8,500 \left( \frac{1.5}{1+q} \right) \left( \frac{M}{10^{8.5} M_\odot} \right)^{1/3} \left( \frac{P}{4.04 \text{ yr}} \right)^{-1/3} \text{ km s}^{-1}$$

or approximately  $0.03c$  for the fiducial parameters chosen in the parentheses on the right ( $q = 0.5$ ,  $M = 10^{8.5} M_\odot$ ,  $P = 4.04 \text{ yr}$ ), where  $M = M_1 + M_2$  is the total binary mass,  $M_{1,2}$  are the individual masses,  $q = M_2/M_1 \leq 1$  is the mass ratio,  $M_\odot$  is the mass of the Sun,  $P$  is the orbital period,  $G$  is the gravitational constant, and  $c$  is the speed of light. The orbital velocity of the higher-mass primary black hole is  $v_1 = qv_2$ . Even if a minidisk has a steady intrinsic rest-frame luminosity, its apparent flux on Earth is modulated by relativistic Doppler beaming. The photon frequencies suffer relativistic Doppler shift by the factor  $D = [\Gamma(1 - \beta_\parallel)]^{-1}$ , where  $\Gamma = (1 - \beta^2)^{-1/2}$  is the Lorentz factor,  $\beta = v/c$  is the three-dimensional velocity  $v$  in units of the speed of light  $c$ , and  $\beta_\parallel = \beta \cos(\varphi) \sin(i)$  is the component of the velocity along the line of sight, with  $i$  and  $\varphi$  the orbital inclination and phase, respectively. Because the photon phase-space density, which is proportional to  $F_\nu/\nu^3$ , is invariant in special relativity, the apparent flux  $F_\nu$  at a fixed observed frequency  $\nu$  is modified from the flux of a stationary source  $F_\nu^0$  to  $F_\nu = D^3 F_{D^{-1}\nu}^0 = D^{3-\alpha} F_\nu^0$ . The last step assumes an intrinsic power-law spectrum  $F_\nu^0 \propto \nu^\alpha$ . To first order in  $v/c$ , this assumption causes a sinusoidal modulation of the apparent flux along the orbit, by a fractional amplitude  $\Delta F_\nu/F_\nu = \pm(3 - \alpha)v \cos(\varphi/c) \sin(i)$ . Although

light-travel time modulations appear at the same order, they are subdominant to the Doppler modulation. This modulation is analogous to periodic modulations from relativistic Doppler boost predicted<sup>16</sup> and observed for extrasolar planets<sup>17,18</sup> and for a double white-dwarf binary<sup>19</sup>, but here it has a much higher amplitude.

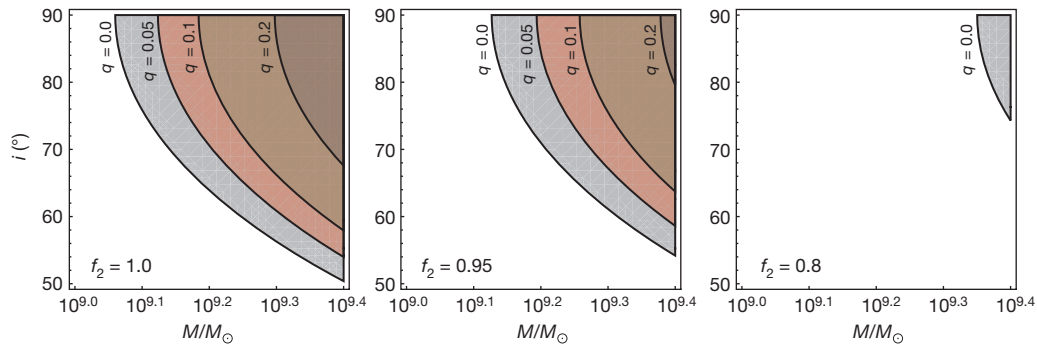
The light curve of PG 1302-102 is well measured over approximately two periods (approximately 10 yr). The amplitude of the variability is  $\pm 0.14 \text{ mag}$  (measured in the optical V band<sup>20</sup>), which corresponds to  $\Delta F_\nu/F_\nu = \pm 0.14$ . The spectrum of PG 1302-102 in and around the V band is well approximated by a double power-law, with  $\alpha \approx 0.7$  (between  $0.50 \mu\text{m}$  and  $0.55 \mu\text{m}$ ) and  $\alpha \approx 1.4$  (between  $0.55 \mu\text{m}$  and  $0.6 \mu\text{m}$ ), except for small deviations caused by broad lines. We obtain an effective single slope  $\alpha_{\text{opt}} = 1.1$  over the entire V band. We conclude that the 14% variability can be attributed to relativistic beaming for a line-of-sight velocity amplitude of  $v \sin(i) = 0.074c = 22,000 \text{ km s}^{-1}$ .

Although large, this velocity can be realized for a massive (high- $M$ ) but unequal-mass (low- $q$ ) binary, whose orbit is viewed not too far from edge-on (high  $\sin(i)$ ). In Fig. 1, we show the required combination of these three parameters that would produce a 0.14-mag variability in the sum-total of Doppler-shifted emission from the primary and the secondary black hole. As this figure shows, the required mass is  $\gtrsim 10^{9.1} M_\odot$ , consistent with the high end of the range that has been inferred for PG 1302-102. The orbital inclination can be in the range  $i = 60^\circ$ – $90^\circ$ . The mass ratio  $q$  has to be low,  $q \lesssim 0.3$ , which is consistent with expectations based on cosmological galaxy merger models<sup>21</sup>, and also with the identification of the optical and binary periods (for  $q \gtrsim 0.3$ , hydrodynamic simulations predict that the mass-accretion rates fluctuate with a period several times longer than the orbital period<sup>22</sup>).

As Fig. 1 shows, fully accounting for the observed optical variability also requires that the bulk of the optical emission arises from gas bound to the faster-moving secondary black hole ( $f_2 \gtrsim 80\%$ ). We find that this condition is naturally satisfied for unequal-mass black holes. Hydrodynamic simulations have shown that for  $0.03 \lesssim q \lesssim 0.1$ , the accretion rate onto the secondary black hole is a factor of 10–20 higher than that onto the primary<sup>13</sup>. Because the secondary captures most of the accreting gas from the circumbinary disk, the primary is ‘starved’, and radiates with a much lower efficiency. In the  $(M, q)$  ranges favoured by the beaming scenario, we find that the primary contributes less than 1% to the total luminosity, and the circumbinary disk contributes less than 20%, leaving the secondary as the dominant source of emission in the three-component system (see Methods).

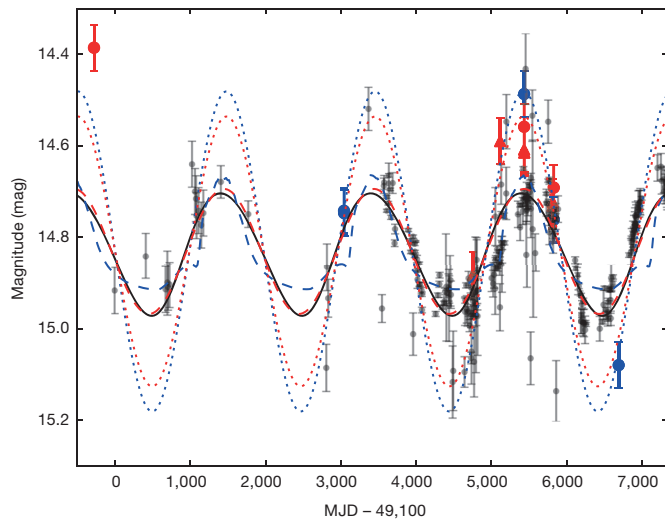
The optical light curve of PG 1302-102 appears remarkably sinusoidal compared to that of the best-studied previous quasi-periodic quasar binary black-hole candidate, which shows periodic bursts<sup>4</sup>. Nevertheless, the light-curve shape deviates from a pure sinusoid. To see if such deviations naturally arise within our model, we maximized the Bayesian likelihood over five parameters (period  $P$ , velocity amplitude  $K$ , eccentricity  $e$ , argument of pericentre  $\omega$ , and an arbitrary reference time  $t_0$ ) of a Kepler orbit<sup>23</sup> and fitted the observed optical light curve. In this procedure, we accounted for additional stochastic physical variability with a broken-power-law power spectrum

<sup>1</sup>Department of Astronomy, Columbia University, 550 West 120th Street, New York, New York 10027, USA.



**Figure 1 | Binary parameters producing the optical flux variations of PG 1302-102 by relativistic boost.** Combinations of total binary mass  $M$ , mass ratio  $q = M_2/M_1$ , and inclination  $i$  that cause  $>13.5\%$  flux variability (or line-of-sight velocity amplitude  $(v/c)\sin(i) \geq 0.07$ ) in the emission from the primary and secondary black holes, computed from the Doppler factor  $D^{3-\alpha}$  with the effective spectral slope of  $\alpha_{\text{opt}} = 1.1$  in the V band. The solid lines

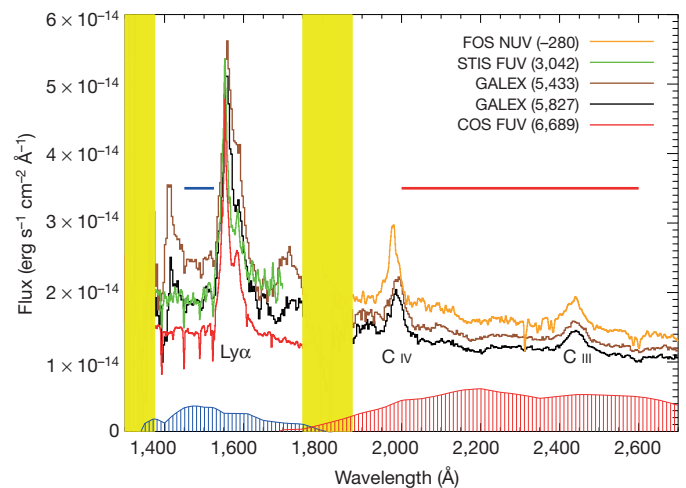
(a ‘damped random walk’<sup>24</sup>) described by two additional parameters. This analysis returns a best-fit with a non-zero eccentricity of  $e = 0.09^{+0.07}_{-0.06}$ , although a Bayesian criterion does not favour this model over a pure sinusoid with fewer parameters (see Methods). We considered an alternative model to explain optical variability of PG 1302-102, in which the luminosity variations track the fluctuations in the mass-accretion rate that is predicted in hydrodynamic simulations<sup>9–11,13,25</sup>. However, the amplitude of these hydrodynamic fluctuations are large (order one), and their shape is ‘bursty’ rather than sinusoid-like<sup>11,13,15</sup>; as a result, we find that they provide a poorer fit to the observations (see Fig. 2 and Methods). Furthermore, for mass ratios  $q \gtrsim 0.05$ , hydrodynamic simulations predict a characteristic pattern of periodicities at multiple frequencies, but an analysis of the periodogram of PG 1302-102 has not uncovered evidence for multiple peaks<sup>26</sup>.



**Figure 2 | The optical and ultraviolet light curves of PG 1302-102.** The grey filled circles with  $1\sigma$  errors are the optical data<sup>6</sup>, superimposed with a best-fit sinusoid (red dashed curve). The solid black curve is the best-fit relativistic light-curve. The blue dashed curve is the best-fit model that was obtained by scaling the mass-accretion rate determined from a hydrodynamic simulation of an unequal-mass ( $q = 0.1$ ) binary<sup>11</sup>. The red and blue filled circles with  $1\sigma$  errors correspond to archival NUV (red) and FUV (blue) spectral observations; the red filled triangles (with  $1\sigma$  errors) represent archival photometric NUV data (see Fig. 3). The UV data include an arbitrary overall normalization to match the mean optical brightness. The red and blue dotted curves are the best-fit relativistic optical light curves with amplitudes scaled up by factors of 2.17 and 2.57, which best match the NUV and FUV data, respectively. MJD, modified Julian day.

correspond to different values of  $q$  as labelled; the shaded regions correspond to intermediate values. We assume that a fraction  $f_2 = 1.0, 0.95$ , or  $0.8$  of the total luminosity arises from the secondary black hole; these values are consistent with fractions found in hydrodynamic simulations<sup>13</sup> (see Methods). The inclination angle is defined such that  $i = 0^\circ$  corresponds to a face-on view of PG 1302-102, and  $i = 90^\circ$  corresponds to an edge-on view.

A simple observational test of relativistic beaming is possible, owing to the strong frequency dependence of the spectral slope of PG 1302-102:  $\alpha = d\ln(F_\nu)/d\ln(\nu)$ . The continuum spectrum of PG 1302-102 is nearly flat with a slope  $\beta_{\text{FUV}} \equiv d\ln(F_\lambda)/d\ln(\lambda) = 0$  in the far-ultraviolet (FUV;  $0.145\text{--}0.1525\ \mu\text{m}$ ) band, where  $F_\lambda$  is the apparent flux at an observed wavelength  $\lambda$ , and shows a tilt with  $\beta_{\text{NUV}} = -0.95$  in the near-ultraviolet (NUV;  $0.20\text{--}0.26\ \mu\text{m}$ ) range; see Fig. 3 and Methods. These slopes translate to  $\alpha_{\text{FUV}} = -2$  and  $\alpha_{\text{NUV}} = -1.05$  in the respective bands, compared to  $\alpha_{\text{opt}} = 1.1$  in the optical. The UV emission can be attributed to the same minidisks that are responsible for the optical light, and would therefore share the same Doppler shifts in frequency. These Doppler shifts would translate into UV variability that is larger by a factor of  $(3 - \alpha_{\text{FUV}})/(3 - \alpha_{\text{opt}}) = 5/1.9 = 2.63$  and  $(3 - \alpha_{\text{NUV}})/(3 - \alpha_{\text{opt}}) = 4.05/1.9 = 2.13$  compared to the optical, and reaches maximum amplitudes of  $\pm 37\%$  (FUV) and  $\pm 30\%$  (NUV).



**Figure 3 | Archival UV spectra of PG 1302-102 from 1992–2011.** FUV and NUV spectra obtained by instruments on the HST and by GALEX, as labelled. COS, cosmic origins spectrograph; FOS, faint object spectrograph; STIS, space telescope imaging spectrograph. Numbers in brackets are the dates (in MJD – 49,100) the data were collected. Vertical yellow bands mark regions outside the spectroscopic range of both GALEX and the HST and contain no useful spectral data. Assignments of the main peaks are given. Ly $\alpha$ , Lyman  $\alpha$ . From each spectrum, average flux measurements (shown in Fig. 2) were computed in one or both of the UV bands over the frequency range indicated by the horizontal bars. The full GALEX photometric band shapes for FUV and NUV photometry are shown for reference as shaded blue and red curves, respectively. Additional GALEX NUV photometric data were also used in Fig. 2. The UV spectra show an offset by as much as  $\pm 30\%$  relative to one another, close to the value expected from relativistic boost (see Methods).



Five separate UV spectra of PG 1302-102 have been collected between 1992 and 2011, by instruments on the Hubble Space Telescope (HST) and on the Galaxy Evolution Explorer (GALEX) satellite (see Fig. 3); additional photometric observations were taken with GALEX at four different times between 2006 and 2009 (shown in Fig. 2). The brightness variations in both the FUV and NUV bands show variability that resembles the optical variability, but with a larger amplitude. Adopting the parameters of our best-fit sinusoid model, and allowing only the amplitude to vary, we find that the UV data yields best-fit variability amplitudes of  $\Delta F_{\text{FUV}}/F_{\text{FUV}} = \pm(35.0 \pm 3.9)\%$  and  $\Delta F_{\text{NUV}}/F_{\text{NUV}} = \pm(29.5 \pm 2.4)\%$  (shown in Fig. 2). These amplitudes are factors of  $2.57 \pm 0.28$  and  $2.17 \pm 0.17$  higher than in the optical, in excellent agreement with the values 2.63 and 2.13 that are expected from the corresponding spectral slopes.

Relativistic beaming provides a simple and robust explanation of the optical periodicity of PG 1302-102. The prediction that the larger UV variations should track the optical light curve can be tested rigorously in the future with measurements of the optical and UV brightness that are collected at or near the same time, are repeated two or more times, are separated by a few months to about 2 yr, and cover up to half of the optical period. A positive result will constitute the first detection of relativistic massive black-hole binary motion; it will also serve as a confirmation of the binary nature of PG 1302-102, remove the ambiguity in the orbital period, and tightly constrain the binary parameters to be close to those shown in Fig. 1.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 10 February; accepted 27 July 2015.**

- Kormendy, J. & Ho, L. C. Coevolution (or not) of supermassive black holes and host galaxies. *Annu. Rev. Astron. Astrophys.* **51**, 511–553 (2013).
- Begelman, M. C., Blandford, R. D. & Rees, M. J. Massive black hole binaries in active galactic nuclei. *Nature* **287**, 307–309 (1980).
- Komossa, S. Observational evidence for binary black holes and active double nuclei. *Mem. Soc. Astron. Ital.* **77**, 733–741 (2006).
- Valtonen, M. J. *et al.* A massive binary black-hole system in OJ 287 and a test of general relativity. *Nature* **452**, 851–853 (2008).
- Liu, T. *et al.* A periodically varying luminous quasar at  $z = 2$  from the Pan-STARRS1 Medium Deep Survey: a candidate supermassive black hole binary in the gravitational wave-driven regime. *Astrophys. J.* **803**, L16 (2015).
- Graham, M. J. *et al.* A possible close supermassive black-hole binary in a quasar with optical periodicity. *Nature* **518**, 74–76 (2015).
- Milosavljević, M. & Merritt, D. in *The Astrophysics of Gravitational Wave Sources*, AIP Conf. Proc. (eds Centrella, J. M. & Barnes, S.) **686**, 201–210 (AIP, 2003).
- Hayasaki, K., Mineshige, S. & Ho, L. C. A supermassive binary black hole with triple disks. *Astrophys. J.* **682**, 1134–1140 (2008).
- Shi, J.-M., Krolik, J. H., Lubow, S. H. & Hawley, J. F. Three-dimensional magnetohydrodynamic simulations of circumbinary accretion disks: disk structures and angular momentum transport. *Astrophys. J.* **749**, 118 (2012).
- Roedig, C. *et al.* Evolution of binary black holes in self gravitating discs. Dissecting the torques. *Astron. Astrophys.* **545**, A127 (2012).
- D'Orazio, D. J., Haiman, Z. & MacFadyen, A. Accretion into the central cavity of a circumbinary disc. *Mon. Not. R. Astron. Soc.* **436**, 2997–3020 (2013).
- Nixon, C., King, A. & Price, D. Tearing up the disc: misaligned accretion on to a binary. *Mon. Not. R. Astron. Soc.* **434**, 1946–1954 (2013).
- Farris, B. D., Duffell, P., MacFadyen, A. I. & Haiman, Z. Binary black hole accretion from a circumbinary disk: gas dynamics inside the central cavity. *Astrophys. J.* **783**, 134 (2014).
- Dunhill, A. C., Cuadra, J. & Dougados, C. Precession and accretion in circumbinary discs: the case of HD 104237. *Mon. Not. R. Astron. Soc.* **448**, 3545–3554 (2015).
- Shi, J.-M. & Krolik, J. H. Three-dimensional MHD simulation of circumbinary accretion disks. II. Net accretion rate. *Astrophys. J.* **807**, 131 (2015).
- Loeb, A. & Gaudi, B. S. Periodic flux variability of stars due to the reflex Doppler effect induced by planetary companions. *Astrophys. J.* **588**, L117–L120 (2003).
- van Kerkwijk, M. H. *et al.* Observations of Doppler boosting in Kepler light curves. *Astrophys. J.* **715**, 51–58 (2010).
- Mazeh, T. & Faigler, S. Detection of the ellipsoidal and the relativistic beaming effects in the CoRoT-3 lightcurve. *Astron. Astrophys.* **521**, L59 (2010).
- Shporer, A. *et al.* A ground-based measurement of the relativistic beaming effect in a detached double white dwarf binary. *Astrophys. J.* **725**, L200–L204 (2010).
- Djorgovski, S. G. *et al.* Exploring the variable sky with the Catalina Real-time Transient Survey. In *The First Year of MAXI: Monitoring Variable X-ray Sources* (eds Mihara, T. & Serino, M.) 32 (MAXI, 2010).
- Volonteri, M., Haardt, F. & Madau, P. The assembly and merging history of supermassive black holes in hierarchical models of galaxy formation. *Astrophys. J.* **582**, 559–573 (2003).
- D'Orazio, D. J., Haiman, Z., Duffell, P., Farris, B. D. & MacFadyen, A. I. A reduced orbital period for the supermassive black hole binary candidate in the quasar PG 1302-102? *Mon. Not. R. Astron. Soc.* **452**, 2540–2545 (2015).
- Wright, J. T. & Gaudi, B. S. in *Planets, Stars and Stellar Systems* Vol. 3 (eds Oswalt, T. D. *et al.*) 489–540 (Springer, 2013).
- Kelly, B. C., Bechtold, J. & Siemiginowska, A. Are the variations in quasar optical flux driven by thermal fluctuations? *Astrophys. J.* **698**, 895–910 (2009).
- MacFadyen, A. I. & Milosavljević, M. An eccentric circumbinary accretion disk and the detection of binary massive black holes. *Astrophys. J.* **672**, 83–93 (2008).
- Charisi, M., Bartos, I., Haiman, Z., Price-Whelan, A. & Márka, S. Multiple periods in the variability of the supermassive black hole binary candidate quasar PG1302-102? *Mon. Not. R. Astron. Soc. Lett.* (in the press).

**Acknowledgements** The authors thank M. Graham, J. Halpern, A. Price-Whelan, J. Andrews, M. Charisi, E. Quataert, and B. Kocsis for discussions. We also thank M. Graham for providing the optical data in electronic form. This work was supported by the National Science Foundation Graduate Research Fellowship under grant no. DGE1144155 (D.J.D.) and by the NASA grant NNX11AE05G (Z.H.).

**Author Contributions** Z.H. conceived and supervised the project, performed the orbital velocity calculations, and wrote the first draft of the paper. D.J.D. computed the emission models and performed the fits to the observed light curve. D.S. analysed the archival UV data. All authors contributed to the text.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to Z.H. (zoltan@astro.columbia.edu).

## METHODS

**V-band emission from a three-component system in PG 1302-102.** Here we assume that the PG 1302-102 supermassive black-hole (SMBH) binary system includes three distinct luminous components: a circumbinary disk (CBD), as well as actively accreting primary and secondary SMBHs. The optical brightness of each of the three components can be estimated once their accretion rates and the black-hole masses  $M_1$  and  $M_2$  are specified. Using the absolute V-band magnitude of PG 1302-102,  $M_V = -25.81$ , and applying a bolometric correction  $BC \approx 10$  (ref. 27), we infer a total bolometric luminosity of  $L_{\text{bol}} = 6.5(BC/10) \times 10^{46} \text{ erg s}^{-1}$ . Bright quasars with the most massive SMBHs ( $M \gtrsim 10^9 M_\odot$ ), have a typical radiative efficiency of  $\epsilon = 0.3$  (ref. 28). Adopting this value, the implied accretion rate is  $\dot{M} = L_{\text{bol}}/(\epsilon c^2) = 3.7 M_\odot \text{ yr}^{-1}$  (where the overdot indicates differentiation with respect to time).

We identify this as the total accretion rate through the CBD, and require that at small radii, the rate is split between the two black holes  $\dot{M} = \dot{M}_1 + \dot{M}_2$  with the ratio  $\eta \equiv \dot{M}_2/\dot{M}_1$ . Hydrodynamic simulations<sup>13</sup> have found that the secondary black hole captures the large majority of the gas, with  $10 \lesssim \eta \lesssim 20$  for  $0.03 \lesssim q \lesssim 0.1$  (where  $q \equiv M_2/M_1$ ). Defining the Eddington ratio of the  $i$ th disk as its accretion rate scaled by its Eddington-limited rate  $f_{i,\text{Edd}} \equiv \dot{M}_i + \dot{M}_{i,\text{Edd}}$  with  $\dot{M}_{\text{Edd}} \equiv L_{\text{Edd}}/(0.1 c^2)$  (here  $L_{\text{Edd}}$  is the Eddington luminosity for the  $i$ th black hole, and we have adopted the fiducial radiative efficiency of  $\epsilon = 0.1$  to be consistent with the standard definition in the literature), we have

$$\begin{aligned} f_{\text{CBD,Edd}} &\approx 0.068 \left( \frac{M}{10^{9.4} M_\odot} \right)^{-1} \\ f_{1,\text{Edd}} &= f_{\text{CBD,Edd}} \frac{1+q}{1+\eta} \approx 0.0034 \left( \frac{f_{\text{CBD,Edd}}}{0.068} \right) \left( \frac{1+q}{1.05} \right) \left( \frac{21}{1+\eta} \right) \\ f_{2,\text{Edd}} &= \eta \frac{f_{1,\text{Edd}}}{q} \approx 1.37 \left( \frac{f_{1,\text{Edd}}}{0.0034} \right) \left( \frac{\eta}{20} \right) \left( \frac{0.05}{q} \right) \end{aligned}$$

where the subscripts 1 and 2 refer to the primary and the secondary black holes, respectively. We adopt a standard, radiatively efficient, geometrically thin, optically thick Shakura–Sunyaev disk model<sup>29</sup> to compute the luminosities produced in the CBD and the circumsecondary disk (CSD). Although the secondary black hole is accreting at a super-Eddington rate, recent three-dimensional radiation magnetohydrodynamic simulations of super-Eddington accretion find radiative efficiencies comparable to the values in standard thin-disk models<sup>30</sup>. On the other hand, the primary black hole is accreting below the critical rate  $\dot{M}_1 \lesssim \dot{M}_{\text{ADAF}} \approx 0.027(\alpha/0.3)^2 \dot{M}_{\text{Edd}}$  (ADAF, advection-dominated accretion flow;  $\alpha$  is the viscosity parameter) at which advection dominates the energy balance<sup>31</sup>. We therefore estimate the luminosity of the primary black hole from a radiatively inefficient ADAF<sup>32,33</sup>, rather than from a Shakura–Sunyaev disk. This interpretation is supported by the fact that PG 1302-102 is known to be an extended radio source, with evidence for a jet and bends in the extended radio structure<sup>34</sup>, features that are commonly associated with sub-Eddington sources<sup>35</sup>.

For the radiatively efficient CBD and CSD, the frequency-dependent luminosity is determined by integrating the local, modified black-body flux over the area of the disk:

$$L_\nu = 2\pi \int_{R_{\text{in}}}^{R_{\text{out}}} F_\nu [T_p(r)] r dr$$

where

$$F_\nu = \pi \frac{2\epsilon_\nu^{1/2}}{1+\epsilon_\nu^{1/2}} B_\nu, \quad \epsilon_\nu = \frac{\kappa_\nu^{\text{abs}}}{\kappa_\nu^{\text{abs}} + \kappa_\nu^{\text{es}}}$$

$B_\nu$  is the Planck function,  $\kappa_\nu^{\text{abs}}$  is the frequency-dependent absorption opacity,  $\kappa_\nu^{\text{es}}$  is the electron scattering opacity, and  $r \in (R_{\text{in}}, R_{\text{out}})$  is the radial coordinate, with  $R_{\text{in,out}}$  the inner and outer radii of the appropriate disk. We compute the radial disk-photosphere temperature profile  $T_p$  by equating the viscous heating rate with the modified black-body flux:

$$\frac{3GM\dot{M}}{8\pi r^3} \left[ 1 - \left( \frac{r_{\text{ISCO}}}{r} \right)^{1/2} \right] = \zeta(v, T_p) \sigma T_p^4$$

where

$$\zeta(v, T_p) = \frac{15}{\pi^4} \int \frac{2\epsilon_\nu^{1/2}(x)}{1+\epsilon_\nu^{1/2}(x)} \frac{x^3 e^{-x}}{1-e^{-x}} dx, \quad x \equiv \frac{h\nu}{k_B T_p}$$

$\sigma$  is the Stefan–Boltzmann constant,  $k_B$  is the Boltzmann constant,  $h$  is the Planck constant, and  $r_{\text{ISCO}}$  is the radius of the innermost stable circular orbit (ISCO). When solving for the photosphere temperature, we work in the limit that  $\kappa_\nu^{\text{abs}} \ll \kappa_\nu^{\text{es}}$  following appendix A of ref. 36, and we adopt  $r_{\text{ISCO}} = 6GM_i/c^2$ , and

$(R_{\text{in}}, R_{\text{out}}) = (2a, 200a)$  and  $(R_{\text{in}}, R_{\text{out}}) = (r_{\text{ISCO}}, a(q/3)^{1/3})$  for the CBD and CSD, respectively. Here the superscript  $i$  refers to the  $i$ th disk,  $a$  is the binary separation,  $6GM$  is the location of the ISCO for a Schwarzschild black hole (our results are insensitive to this choice) and  $a(q/3)^{1/3}$  is the Hill radius of the secondary black hole (which provides an upper limit on the size of the CSD<sup>37</sup>).

The optical luminosity of an ADAF is sensitive to the assumed microphysical parameters and its computation is more complicated than that for a thin disk. Here we first compute a reference thin-disk luminosity  $L_{\text{SS}}$  (SS, Shakura–Sunyaev) for the primary black hole, and multiply it by the ratio of the bolometric luminosity of an ADAF to an equivalent thin-disk luminosity from ref. 32:

$$\frac{L_{\text{ADAF}}}{L_{\text{SS}}} \approx 0.008 \left( \frac{\dot{M}/\dot{M}_{\text{Edd}}}{0.0034} \right) \left( \frac{\alpha}{0.3} \right)^{-2}$$

For calculating the reference  $L_{\text{SS}}$ , we adopted parameters that are consistent with ref. 32, in particular,  $\epsilon = 0.1$ . Although the above ratio is for bolometric luminosities, we find that it agrees well with the factor-of-100 difference in the V band shown in figure 6 of ref. 33, between ADAF and thin-disk spectra, with parameters similar to PG 1302-102 ( $10^9 M_\odot$ ,  $\dot{M} = \dot{M}_{\text{ADAF}} = 10^{-1.5} \dot{M}_{\text{Edd}}$ ,  $\alpha \approx 0.3$ ).

Extended Data Figure 1 shows the thin-disk CBD and CSD spectra for a total Eddington ratio of  $f_{\text{CBD,Edd}} = 0.07$ , consistent with the high-mass estimates for PG 1302-102 that are needed for the beaming scenario ( $M = 10^{9.4} M_\odot$  and  $q = 0.05$ ). The red dot shows the reduced V-band luminosity of an ADAF onto the primary. The secondary clearly dominates the total V-band luminosity, with the primary contributing less than 1%, and the CBD contributing approximately 14%. In practice, the contribution from the CBD becomes non-negligible only for the smallest binary masses and lowest mass ratios (reaching 20% for  $M < 10^9 M_\odot$  and  $q < 0.025$ ).

We compute the contributions of each of the three components to the total luminosity,  $L_{\text{tot}}^V = L_1^V + L_2^V + L_{\text{CBD}}^V$ , and the corresponding total fractional-modulation amplitude  $\Delta L_{\text{tot}}^V/L_{\text{tot}}^V = (\Delta L_1^V + \Delta L_2^V)/L_{\text{tot}}^V$ , for each value of the total mass  $M$  and mass ratio  $q$ . The primary is assumed to be Doppler modulated with a line-of-sight velocity  $v_1 = -qv_2$ , whereas the emission from the CBD is assumed to be constant over time ( $\Delta L_{\text{CBD}}^V = 0$ ). Extended Data Figure 2 shows regions in  $(M, q, i)$  parameter space where the total luminosity variation due to relativistic beaming exceeds 14%. This figure recreates Fig. 1 of the main text, but using the luminosity contributions computed self-consistently in the above model, rather than assuming a constant value of  $f_2$ . Because the secondary is found to be dominant, the relativistic-beaming scenario is consistent with a wide range of binary parameters.

**Model fitting to the PG 1302-102 optical light curve.** We fitted models to the observed light curve of PG 1302-102 by maximizing the Bayesian likelihood  $\mathcal{L} \propto \det[\text{Cov}^D \text{Cov}^{\text{ph}}]^{-1/2} \exp(-\chi^2/2)$ , where  $\chi^2 \equiv \mathbf{Y}^T (\text{Cov})^{-1} \mathbf{Y}$  and  $\mathbf{Y} \equiv \mathbf{O} - \mathbf{M}$  is the difference vector between the mean flux predicted in a model and the observed flux at each observation time  $t_i$ . Here  $\text{Cov}$  is the covariance matrix of flux uncertainties, allowing for correlations between fluxes measured at different  $t_i$ . We include two types of uncertainties: (1) random (uncorrelated) measurement errors

$$\text{Cov}_{ij}^{\text{ph}} = \begin{cases} \sigma_i^2 & i=j \\ 0 & i \neq j \end{cases}$$

where  $\sigma_i^2$  is the variance in the photometric measurement for the  $i$ th data point (as reported in ref. 6); and (2) correlated noise due to intrinsic quasar variability, with covariance between the  $i$ th and  $j$ th data points

$$\text{Cov}_{ij}^D = \sigma_D^2 \exp \left[ \frac{-|t_i - t_j|}{(1+z)\tau_D} \right]$$

The parameters  $\sigma_D$  and  $\tau_D$  determine the amplitude and rest-frame coherence time, respectively, of correlated noise described by the damped random walk model<sup>24</sup>, and the factor of  $(1+z)$  converts  $\tau_D$  to the observer's frame where the  $t_i$  are measured. The normalization of the Bayesian likelihood depends on these parameters, and therefore the normalization must be included when maximizing the likelihood over these parameters<sup>38</sup>. The covariance matrix for the total noise is given by  $\text{Cov} = \text{Cov}^D + \text{Cov}^{\text{ph}}$ . We assume both types of noise are Gaussian, which provides a good description of observed quasar variability<sup>39</sup>.

We then fit the following four different types of models to the data.

- (1) A relativistic beaming model with  $5 + 2 = 7$  model parameters: eccentricity  $e$ , argument of pericentre  $\omega$ , amplitude  $K$ , phase  $t_0$ , and orbital period  $P$ , as well as the two noise parameters  $\sigma_D$  and  $\tau_D$ .
- (2) An accretion rate model with  $3 + 2 = 5$  model parameters: amplitude  $K$ , phase  $t_0$ , and period  $P$ , as well as the two noise parameters  $\sigma_D$  and  $\tau_D$ . This model assumes that the light curve of PG 1302-102 tracks the mass-accretion rates that are predicted in hydrodynamic simulations. For near-equal-mass binaries, several

studies have found that the mass-accretion rates fluctuate periodically, but resemble a series of sharp bursts, unlike the smoother, sinusoid-like shape of the light curve of PG 1302-102. To our knowledge, only three studies so far have simulated unequal-mass ( $q \leq 0.1$ ) SMBH binaries<sup>11,13,15</sup>. The accretion rates for these binaries are less 'bursty'; among all of the cases in these three studies, the  $q = 0.075$  and  $q = 0.1$  binaries in ref. 11 resemble the light curve of PG 1302-102 most closely (see Extended Data Fig. 3). Here we adopt the published accretion curve for  $q = 0.1$ , and perform a fit to PG 1302-102 by allowing an arbitrary linear scaling in time and amplitude, as well as a shift in phase; this gives us the three free parameters for this model. (We find that the  $q = 0.075$  case provides a worse fit.) (3) A sinusoid model with  $3 + 2 = 5$  parameters: amplitude  $K$ , phase  $t_0$ , and period  $P$ , as well as the two noise parameters  $\sigma_D$  and  $\tau_D$ . This model is equivalent, to first order in  $v/c$ , to the beaming model restricted to a circular binary orbit. (4) A constant luminosity model with 2 parameters, the noise parameters  $\sigma_D$  and  $\tau_D$ . This model is for reference only, to quantify how poor the fit is with only these parameters.

In each of these models, we fixed the mean flux to correspond to the mean magnitude  $M$  that is inferred from the optical data; allowing the mean to be an additional free parameter did not change our results. The highest maximum likelihood is found for the beaming model, with best-fit values of  $P = 1,996^{+29}_{-35}$  days,  $K = 0.065^{+0.007}_{-0.006}$ ,  $e = 0.09^{+0.07}_{-0.06}$ ,  $\cos(\omega) = -0.65^{+1.2}_{-0.06}$ , and  $t_0 = 718^{+422}_{-34}$  days, where the reference point  $t_0$  is measured from MJD = 49,100. Uncertainties are computed with the 'emcee' code<sup>40</sup>, which implements a Markov chain Monte Carlo algorithm, and which we use to sample the seven-dimensional posterior probability of the model given the data in ref. 6. We use 28 individual chains to sample the posterior for 1,024 steps each. Throwing away the first 600 steps ('burning in'), we run for 424 steps and for each parameter we quote best-fit values corresponding to the maximum posterior probability, with errors given by the 85th and 15th percentile values (marginalized over the other six parameters). The best-fit noise parameters are  $(\sigma_D, \tau_D) = (0.049^{+0.016}_{-0.001}$  mag,  $37.6^{+35.2}_{-1.5}$  days). The best-fit model has a reduced  $\chi^2/(N - 1 - 7) \approx 2.1$ , where  $N = 245$  is the number of data points.

To assess which of the models is favoured by the data, we use the Bayesian information criterion (BIC), a standard method for comparing different models that penalizes models with a larger number of free parameters<sup>41</sup>. Specifically,  $\text{BIC} = -2 \ln(\mathcal{L}) + k \ln(N)$ , where the first term is evaluated using the best-fit parameters in each of the models and  $k$  is the number of model parameters. We find the following differences  $\Delta\text{BIC}$  between pairs of models:

$\text{BIC}_{\text{Acc}} - \text{BIC}_{\text{Beam}} = 4.0$  (the beaming model is preferred over the accretion model);

$\text{BIC}_{\text{Acc}} - \text{BIC}_{\text{Sin}} = 14.9$  (the sinusoid model is strongly preferred over the accretion model);

$\text{BIC}_{\text{Sin}} - \text{BIC}_{\text{Beam}} = -10.9$  (the sinusoid model is strongly preferred over the beaming model);

$\text{BIC}_{\text{Const}} - \text{BIC}_{\text{Beam}} = 11.5$  (the beaming model is strongly preferred over pure noise); and

$\text{BIC}_{\text{Const}} - \text{BIC}_{\text{Sin}} = 22.4$  (the sinusoid model is strongly preferred over pure noise).

We conclude that a sinusoid, or equivalently the beaming model restricted to a circular binary, is the preferred model. This model is very strongly favoured over the best-fit accretion model (see Extended Data Fig. 3), with  $\Delta\text{BIC} > 14.9$ . For the assumed Gaussian distributions, this corresponds to an approximate likelihood ratio of  $\exp(-14.9/2) \approx 5.7 \times 10^{-4}$ . Although our best-fit beaming model has a small non-zero eccentricity, the seven-parameter eccentric model is disfavoured (by  $\Delta\text{BIC} = 11.5$ ) over the five-parameter circular case.

We conservatively allowed the amplitude of accretion-rate fluctuations to be a free parameter in the accretion models, but we note that the accretion-rate variability measured in hydrodynamic simulations exhibits large (order one) deviations from the mean, even for  $0.05 < q < 0.1$  binaries<sup>11,13,15</sup>. In the accretion-rate model, an additional physical mechanism needs to be invoked to damp the fluctuations to the smaller, approximately 14% amplitude seen in PG 1302-102 (such as a more substantial contribution from the CBD and/or the primary).

**Disk precession.** The lowest BIC model, with a steady accretion rate and a relativistic boost from a circular orbit, has a reduced  $\chi^2 = 2.1$ , indicating that the relativistic-boost model with intrinsic noise does not fully describe the observed light curve. The residuals could be explained by a lower-amplitude periodic modulation in the mass-accretion rate, which is expected to have a non-sinusoidal shape (with sharper peaks and broader troughs, as mentioned above<sup>13</sup>). Alternatively, the minidisks, which we implicitly assumed to be co-planar with the binary orbit, could instead have a substantial tilt<sup>12</sup>.

A circumsecondary minidisk that is tilted with respect to the orbital plane of the binary will precess around the binary angular-momentum vector, causing additional photometric variations due to the changing projected area of the disk on the sky. The precession timescale can be estimated from the total angular momentum

of the secondary disk and the torque exerted on it by the primary black hole. The ratio of the precession period to the orbital period of the binary is<sup>42</sup>

$$\frac{P_{\text{prec}}}{P_{\text{orb}}} = -\frac{8}{\sqrt{3}} \frac{\sqrt{1+q}}{\cos(\delta)}$$

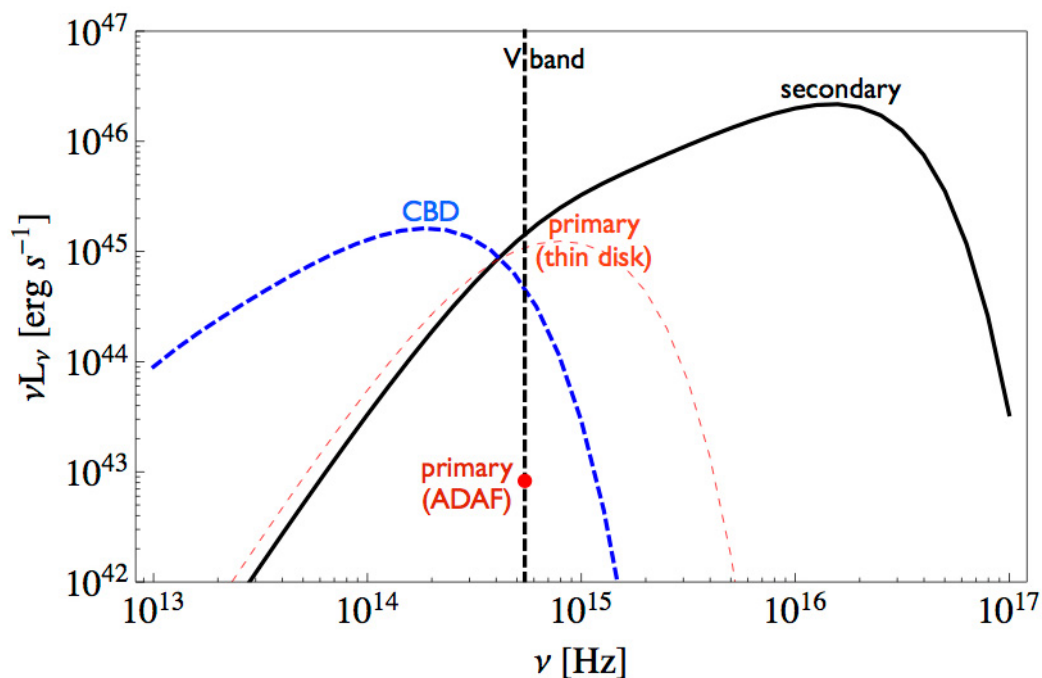
where  $\delta \in (-\pi/2, \pi/2)$  is the angle between the angular-momentum vectors of the disk and the binary, and we have chosen the outer edge of the minidisk to coincide with the Hill sphere of the secondary black hole  $R_{\text{H}} = (q/3)^{1/3} a$ , for binary semi-major axis  $a$ . This choice gives the largest secondary disk and the shortest precession rates. For small binary mass ratios, consistent with the relativistic beaming scenario, the precession can be as short as  $4.8P_{\text{orb}}$ , which causes variations on a timescale that spans the current observations of PG 1302-102. The precession timescale would be longer ( $> 20P_{\text{orb}}$ ) for a smaller secondary disk that is tidally truncated at  $0.27q^{0.3}a$  (ref. 43), and with a more inclined ( $45^\circ$ ) disk.

**Archival UV data.** FUV (0.14–0.175  $\mu\text{m}$ ) and NUV (0.19–0.27  $\mu\text{m}$ ) spectra of PG 1302-102 were obtained by the HST and the GALEX since 1992. HST FOS NUV spectra were obtained on 17 July 1992 (pre-COSTAR)<sup>44</sup>. HST STIS FUV spectra were obtained on 21 August 2001 (ref. 45). GALEX FUV and NUV spectra were obtained on 8 March 2008 and 6 April 2009, and HST COS FUV spectra were obtained on 28 January 2011. All data are publicly available through the Mikulski Archive for Space Telescopes at <http://archive.stsci.edu>. All measurements were spectrophotometrically calibrated, and binned or smoothed to a resolution of 1–3 Å. The spectra (Fig. 2) have errors per bin that are typically less than 2%; published absolute photometric accuracies are better than 5%.

From each spectrum, average flux measurements (Fig. 2) were obtained in one or both of two discrete bands: FUV continuum (0.145–0.1525  $\mu\text{m}$ ; a range chosen to avoid the Ly $\alpha$  line) and NUV continuum (0.20–0.26  $\mu\text{m}$ ). For the GALEX NUV photometric data (also used in Fig. 2) we adopted a small correction (0.005 mag) for the transformation from the GALEX NUV to our NUV continuum band. GALEX FUV photometric data were not used because of the substantial contribution from redshifted Ly $\alpha$ . The broad lines in the UV spectra (in Fig. 3) do not show a large Doppler shift ( $\Delta\lambda = (v/c)\lambda \approx 140$  Å). This is unsurprising, because the broad line widths (2,500–4,500  $\text{km s}^{-1}$ ) are much smaller than the inferred relativistic line-of-sight velocities, and are expected to be produced by gas at larger radii, unrelated to the rapidly orbiting minidisks that produce the featureless thermal continuum emission<sup>22</sup>.

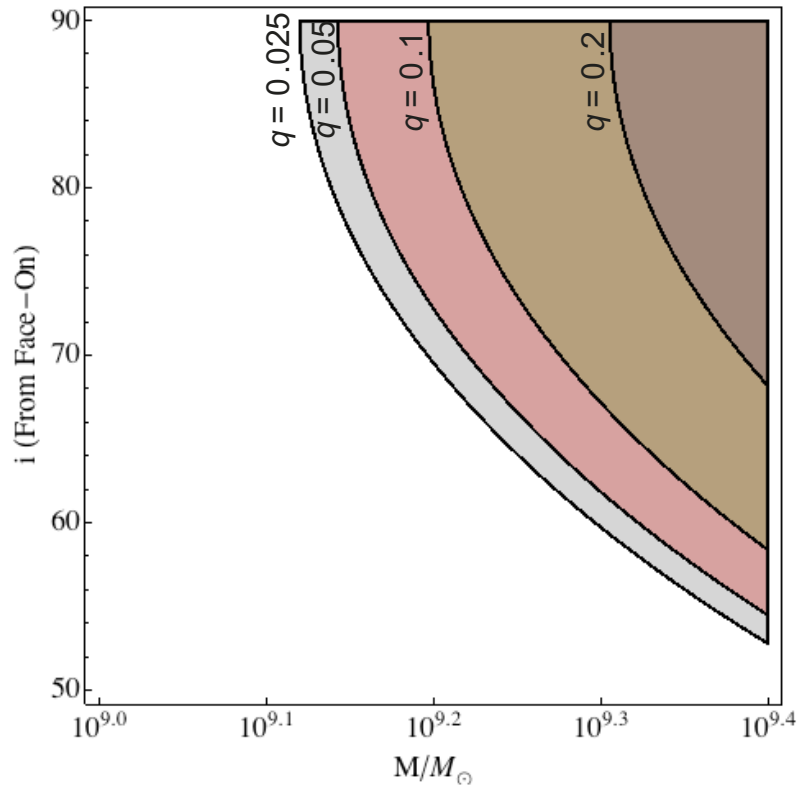
27. Richards, G. T. *et al.* Spectral energy distributions and multiwavelength selection of type 1 quasars. *Astrophys. J.* **166** (Suppl.), 470–497 (2006).
28. Yu, Q. & Tremaine, S. Observational constraints on growth of massive black holes. *Mon. Not. Astron. R. Soc.* **335**, 965–976 (2002).
29. Shakura, N. I. & Sunyaev, R. A. Black holes in binary systems. Observational appearance. *Astron. Astrophys.* **24**, 337–355 (1973).
30. Jiang, Y.-F., Stone, J. M. & Davis, S. W. A global three-dimensional radiation magneto-hydrodynamic simulation of super-Eddington accretion disks. *Astrophys. J.* **796**, 106 (2014).
31. Narayan, R. & McClintock, J. E. Advection-dominated accretion and the black hole event horizon. *New Astron. Rev.* **51**, 733–751 (2008).
32. Mahadevan, R. Scaling laws for advection-dominated flows: applications to low-luminosity galactic nuclei. *Astrophys. J.* **477**, 585–601 (1997).
33. Narayan, R., Mahadevan, R. & Quataert, E. in *Theory of Black Hole Accretion Disks* (eds Abramowicz, M. A. *et al.*) 148–182 (Cambridge Univ. Press, 1998).
34. Hutchings, J. B., Morris, S. C., Gower, A. C. & Lister, M. L. Correlated optical and radio structure in the QSO 1302-102. *Publ. Astron. Soc. Pac.* **106**, 642–645 (1994).
35. Wang, J.-M., Ho, L. C. & Staubert, R. The central engines of radio-loud quasars. *Astron. Astrophys.* **409**, 887–898 (2003).
36. Tanaka, T. & Menou, K. Time-dependent models for the afterglows of massive black hole mergers. *Astrophys. J.* **714**, 404–422 (2010).
37. Artymowicz, P. & Lubow, S. H. Dynamics of binary-disk interaction. 1. Resonances and disk gap sizes. *Astrophys. J.* **421**, 651–667 (1994).
38. Kozłowski, S. *et al.* Quantifying quasar variability as part of a general approach to classifying continuously varying sources. *Astrophys. J.* **708**, 927–945 (2010).
39. Andrae, R., Kim, D.-W. & Bailer-Jones, C. A. L. Assessment of stochastic and deterministic models of 6304 quasar lightcurves from SDSS Stripe 82. *Astron. Astrophys.* **554**, A137 (2013).
40. Foreman-Mackey, D., Hogg, D. W., Lang, D. & Goodman, J. emcee: the MCMC hammer. *Publ. Astron. Soc. Pac.* **125**, 306–312 (2013).
41. Kass, R. E. & Raftery, A. E. Bayes factors. *J. Am. Stat. Assoc.* **90**, 773–795 (1995).
42. Lai, D. Star-disc-binary interactions in protoplanetary disc systems and primordial spin-orbit misalignments. *Mon. Not. R. Astron. Soc.* **440**, 3532–3544 (2014).
43. Roedig, C., Krolik, J. H. & Miller, M. C. Observational signatures of binary supermassive black holes. *Astrophys. J.* **785**, 115 (2014).
44. Evans, I. N. & Koratkar, A. P. A complete atlas of recalibrated *Hubble Space Telescope* Faint Object Spectrograph spectra of active galactic nuclei and auasars. I. Pre-COSTAR spectra. *Astrophys. J.* **150** (Suppl.), 73–164 (2004).
45. Cooksey, K. L., Prochaska, J. X., Chen, H.-W., Mulchaey, J. S. & Weiner, B. J. Characterizing the low-redshift intergalactic medium toward PKS 1302-102. *Astrophys. J.* **676**, 262–285 (2008).





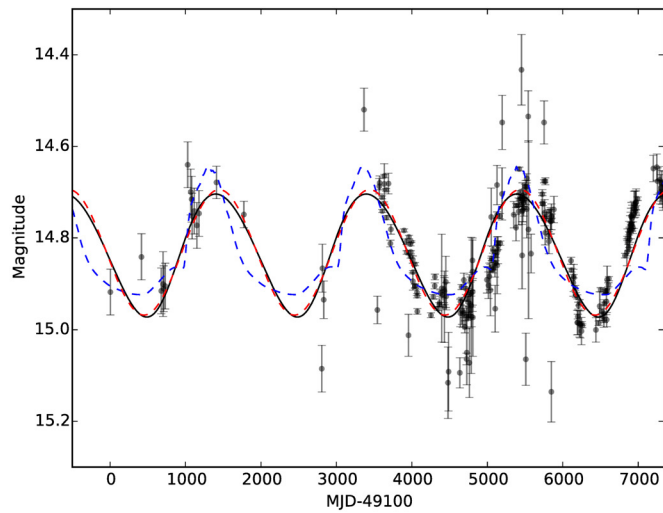
**Extended Data Figure 1 | Model spectrum of PG 1302-102.** Circumbinary (dashed blue) and circumsecondary (solid black) disk spectra for a total binary mass of  $10^{9.4} M_\odot$ , binary mass ratio of  $q = 0.05$ , and ratio of accretion rates  $\dot{M}_2/\dot{M}_1 = 20$ . A vertical dashed line marks the centre of the V band and the

approximate flux from an advection-dominated accretion flow (ADAF) is shown as a red dot for the V-band contribution of the primary. The spectrum for a radiatively efficient, thin disk around the primary is shown by the thin red dashed curve for reference.

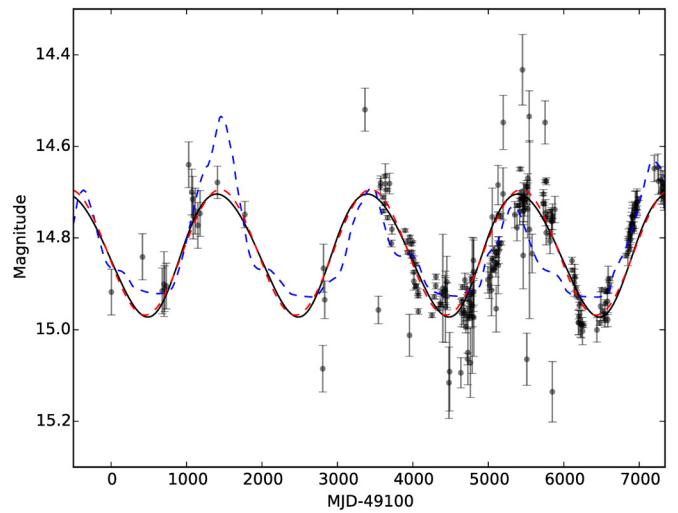


**Extended Data Figure 2 | Parameter combinations for which the combined V-band luminosity of the three-component system varies by the required 0.14 mag.**  $M$  is the binary mass,  $q$  is the mass ratio, and  $i$  is the orbital inclination angle. This figure is analogous to Fig. 1, except instead of adopting an ad-hoc fractional luminosity contribution  $f_2$  by the secondary, the

luminosities of each of the three components are computed from a model: the luminosity of the primary is assumed to arise from an ADAF, whereas the luminosity of the secondary is generated by a modestly super-Eddington thin disk. Emission from the circumbinary disk is also from a thin disk, and is negligible except for binaries with the lowest mass ratio  $q \lesssim 0.01$  (see text).



**Extended Data Figure 3 | Model fits to the optical light curve of PG 1302-102.** Best-fit curves assuming relativistic boost from a circular binary (solid black curves), a pure sinusoid (red dotted curves), and accretion rate variability



adopted from hydrodynamic simulations<sup>11</sup> (blue dashed curves) for a  $q = 0.075$  (left) and a  $q = 0.1$  (right) binary. The grey points with  $1\sigma$  error bars are the data for PG 1302-102 (ref. 6).



# Spawning rings of exceptional points out of Dirac cones

Bo Zhen<sup>1\*</sup>, Chia Wei Hsu<sup>1,2\*</sup>, Yuichi Igarashi<sup>1,3\*</sup>, Ling Lu<sup>1</sup>, Ido Kaminer<sup>1</sup>, Adi Pick<sup>1,4</sup>, Song-Liang Chua<sup>5</sup>, John D. Joannopoulos<sup>1</sup> & Marin Soljačić<sup>1</sup>

The Dirac cone underlies many unique electronic properties of graphene<sup>1</sup> and topological insulators, and its band structure—two conical bands touching at a single point—has also been realized for photons in waveguide arrays<sup>2</sup>, atoms in optical lattices<sup>3</sup>, and through accidental degeneracy<sup>4,5</sup>. Deformation of the Dirac cone often reveals intriguing properties; an example is the quantum Hall effect, where a constant magnetic field breaks the Dirac cone into isolated Landau levels. A seemingly unrelated phenomenon is the exceptional point<sup>6,7</sup>, also known as the parity–time symmetry breaking point<sup>8–11</sup>, where two resonances coincide in both their positions and widths. Exceptional points lead to counter-intuitive phenomena such as loss-induced transparency<sup>12</sup>, unidirectional transmission or reflection<sup>11,13,14</sup>, and lasers with reversed pump dependence<sup>15</sup> or single-mode operation<sup>16,17</sup>. Dirac cones and exceptional points are connected: it was theoretically suggested that certain non-Hermitian perturbations can deform a Dirac cone and spawn a ring of exceptional points<sup>18–20</sup>. Here we experimentally demonstrate such an ‘exceptional ring’ in a photonic crystal slab. Angle-resolved reflection measurements of the photonic crystal slab reveal that the peaks of reflectivity follow the conical band structure of a Dirac cone resulting from accidental degeneracy, whereas the complex eigenvalues of the system are deformed into a two-dimensional flat band enclosed by an exceptional ring. This deformation arises from the dissimilar radiation rates of dipole and quadrupole resonances, which play a role analogous to the loss and gain in parity–time symmetric systems. Our results indicate that the radiation existing in any open system can fundamentally alter its physical properties in ways previously expected only in the presence of material loss and gain.

Closed and lossless physical systems are described by Hermitian operators, which guarantee realness of the eigenvalues and a complete set of eigenfunctions that are orthogonal to each other. On the other hand, systems with open boundaries<sup>7,21</sup> or with material loss and gain<sup>9–17,19</sup> are non-Hermitian<sup>6</sup>, and have non-orthogonal eigenfunctions with complex eigenvalues where the imaginary part corresponds to decay or growth. The most drastic difference between Hermitian and non-Hermitian systems is that the latter exhibit exceptional points (EPs) where both the real and the imaginary parts of the eigenvalues coalesce. At an EP, two (or more) eigenfunctions collapse into one so the eigenspace no longer forms a complete basis, and this eigenfunction becomes orthogonal to itself under the unconjugated ‘inner product’<sup>6,7</sup>. To date, most studies of the EP and its intriguing consequences concern parity–time symmetric systems that rely on material loss and gain<sup>9–17,19</sup>, but EPs are a general property that require only non-Hermiticity. Here, we show the existence of EPs in a photonic crystal slab with negligible absorption loss and no artificial gain. When a Dirac-cone system has dissimilar radiation rates, the band structure is altered abruptly to show branching features with a ring of EPs. We

provide a complete picture of this system, ranging from an analytic model and numerical simulations to experimental observations; taken together, these results illustrate the role of radiation-induced non-Hermiticity that bridges the study of EPs and the study of Dirac cones.

We start by showing that non-Hermiticity from radiation can deform an accidental Dirac point into a ring of EPs. First, consider a two-dimensional photonic crystal (Fig. 1a inset), where a square lattice (periodicity  $a$ ) of circular air holes (radius  $r$ ) is introduced in a dielectric material. This is a Hermitian system, as there is no material gain or loss and no open boundary for radiation. By tuning a system parameter (for example,  $r$ ), one can achieve accidental degeneracy between a quadrupole mode and two degenerate dipole modes at the  $\Gamma$  point (centre of the Brillouin zone), leading to a linear Dirac dispersion due to the anti-crossing between two bands with the same symmetry<sup>4,22</sup>. The accidental Dirac dispersion from the effective Hamiltonian model (see equation (1) below with  $\gamma_d = 0$ ) is shown as solid lines in Fig. 1a, agreeing with numerical simulation results (symbols). In the effective Hamiltonian we do not consider the dispersionless third band (grey line) owing to symmetry arguments (Supplementary Information section I), although this third band cannot be neglected in certain calculations, including the Berry phase and effective medium properties<sup>23</sup>.

Next, we consider a similar, but open, system: a photonic crystal slab (Fig. 1b inset) with finite thickness  $h$ . With the open boundary, modes within the radiation continuum become resonances because they radiate by coupling to extended plane waves in the surrounding medium. Non-Hermitian perturbations need to be included in the Hamiltonian to account for the radiation loss. To the leading order, radiation of the dipole mode can be described by adding an imaginary part  $-i\gamma_d$  to the Hamiltonian, while the quadrupole mode does not radiate owing to its symmetry mismatch with the plane waves<sup>24</sup>. Specifically, at the  $\Gamma$  point the system has  $C_2$  rotational symmetry (invariant under  $180^\circ$  rotation around the  $z$  axis), and the quadrupole mode does not couple to the radiating plane wave because the former has a field profile  $\mathbf{E}(\mathbf{r})$  that is even under  $C_2$  rotation,  $\mathbf{E}(\mathbf{r}) = \hat{O}_{C_2}\mathbf{E}(\mathbf{r})$ , whereas the latter is odd,  $\mathbf{E}(\mathbf{r}) = -\hat{O}_{C_2}\mathbf{E}(\mathbf{r})$ . The effective Hamiltonian is

$$H_{\text{eff}} = \begin{pmatrix} \omega_0 & v_g |\mathbf{k}| \\ v_g |\mathbf{k}| & \omega_0 - i\gamma_d \end{pmatrix} \quad (1)$$

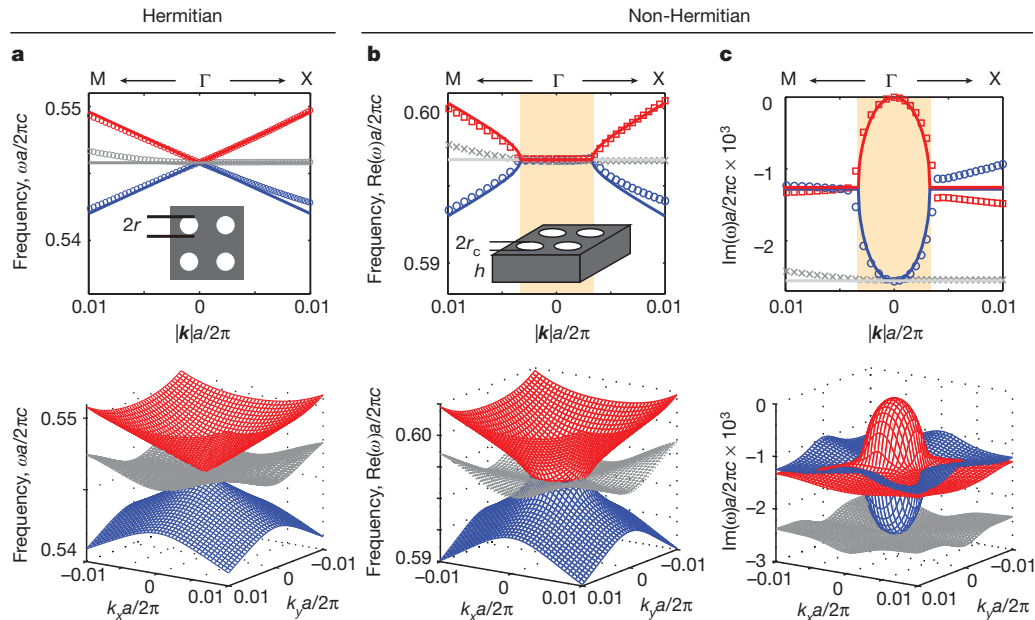
with complex eigenvalues

$$\omega_{\pm} = \omega_0 - i\frac{\gamma_d}{2} \pm v_g \sqrt{|\mathbf{k}|^2 - k_c^2} \quad (2)$$

where  $\omega_0$  is the frequency at accidental degeneracy,  $v_g$  is the group velocity of the linear Dirac dispersion in the absence of radiation,  $|\mathbf{k}|$  is the magnitude of the in-plane wavevector ( $k_x, k_y$ ), and  $k_c \equiv \gamma_d/2v_g$ . Here, one of the three bands is decoupled from the other two and is not included in equation (1) (see Supplementary Information section II). In equation (2), a ring defined by  $|\mathbf{k}| = k_c$  separates the  $k$  space

<sup>1</sup>Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. <sup>2</sup>Department of Applied Physics, Yale University, New Haven, Connecticut 06520, USA. <sup>3</sup>Smart Energy Research Laboratories, NEC Corporation, 34 Miyuiga-ka, Tsukuba, Ibaraki 305-8501, Japan. <sup>4</sup>Department of Physics, Harvard University, Cambridge, Massachusetts 02138, USA. <sup>5</sup>DSO National Laboratories, 20 Science Park Drive, Singapore 118230, Singapore.

\*These authors contributed equally to this work.



**Figure 1 | Accidental degeneracy in Hermitian and non-Hermitian photonic crystals.** **a**, Band structure of a two-dimensional photonic crystal consisting of a square lattice of circular air holes. Tuning the radius  $r$  leads to accidental degeneracy between a quadrupole band and two doubly degenerate dipole bands, resulting in two bands with linear Dirac dispersion (red and blue) and a flat band (grey). **b**, **c**, The real (**b**) and imaginary (**c**) parts of the eigenvalues of an open, and therefore non-Hermitian, system: a photonic crystal slab with finite thickness,  $h$ . By tuning the radius, accidental degeneracy in the real part can be achieved, but the Dirac dispersion is deformed owing to the non-Hermiticity. The analytic model predicts that the

real (imaginary) part of the eigenvalue stays as a constant inside (outside) a ring in the wavevector space, indicating two flat bands in dispersion, with a ring of exceptional points (EPs) where both the real and the imaginary parts are degenerate. The orange shaded regions correspond to the inside of the ring. In the upper panels of **a–c**, solid lines are predictions from the analytic model and symbols are from numerical simulations: red squares represent the band connecting to the quadrupole mode at the centre; blue circles represent the band connecting to the dipole mode at the centre; and grey crosses represent the third band that is decoupled from the previous two due to symmetry. The three-dimensional plots in the lower panels are from simulations.

into two regions: inside the ring ( $|\mathbf{k}| < k_c$ ),  $\text{Re}(\omega_{\pm})$  are dispersionless and degenerate; outside the ring ( $|\mathbf{k}| > k_c$ ),  $\text{Im}(\omega_{\pm})$  are dispersionless and degenerate. In the vicinity of  $k_c$ ,  $\text{Im}(\omega_{\pm})$  and  $\text{Re}(\omega_{\pm})$  exhibit square-root dispersion (also known as branching behaviour<sup>6</sup>) inside and outside the ring, respectively. Exactly on the ring ( $|\mathbf{k}| = k_c$ ), the two eigenvalues  $\omega_{\pm}$  are degenerate in both real and imaginary parts; meanwhile, the matrix  $H_{\text{eff}}$  becomes defective with an incomplete eigenspace spanned by only one eigenvector  $(1, -i)^T$  that is orthogonal to itself under the unconjugated ‘inner product’, given by  $\mathbf{a}^T \mathbf{b}$  for vectors  $\mathbf{a}$  and  $\mathbf{b}$ . This self-orthogonality is the definition of EPs; hence, here we have not just one EP, but a continuous ring of EPs. We call it an exceptional ring.

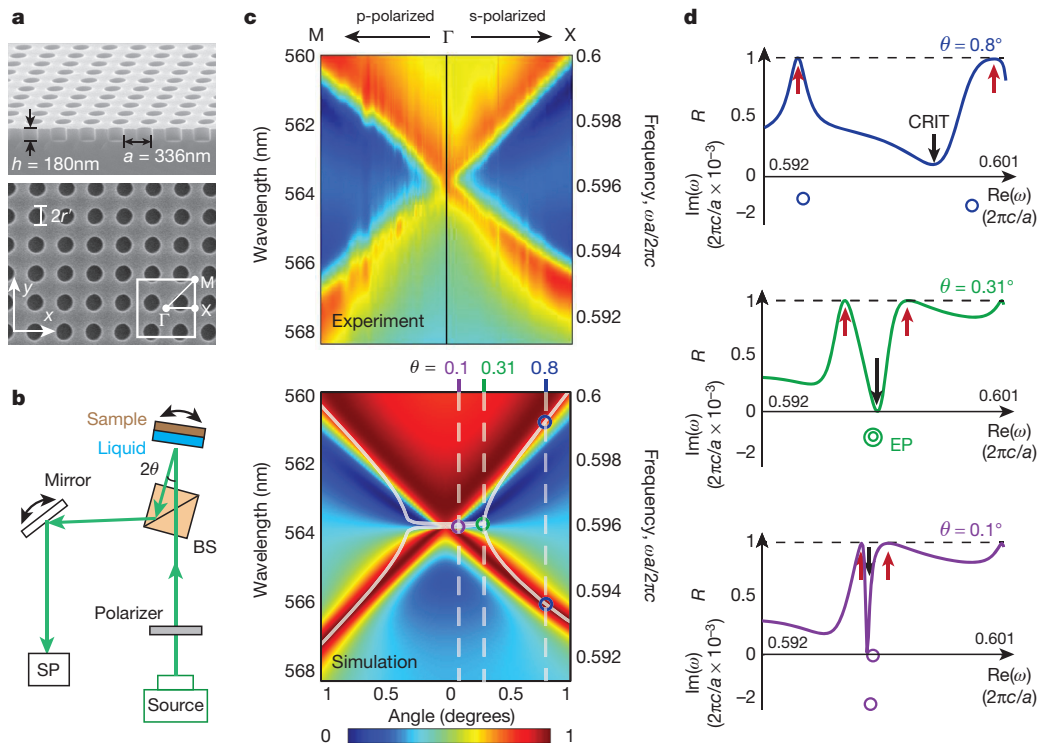
Figure 1b, c shows the complex eigenvalues of the photonic crystal slab structure calculated numerically (symbols), which closely follow the analytic model of equation (2) shown as solid lines in the figure. In Supplementary Fig. 1, we show that the two eigenvectors indeed coalesce into one at the EP, which is impossible in Hermitian systems (also see Supplementary Information section III). When the radius  $r$  of the holes is tuned away from accidental degeneracy, the exceptional ring and the associated branching behaviour disappear, as shown in Supplementary Fig. 2. Several properties of the photonic crystal slab contribute to the existence of this exceptional ring. Owing to periodicity, one can probe the dispersion from two degrees of freedom,  $k_x$  and  $k_y$ , in just one structure. The open boundary provides radiation loss, and the  $C_2$  rotational symmetry differentiates the radiation loss of the dipole mode and of the quadrupole mode.

We can rigorously show that the exceptional ring exists in realistic photonic crystal slabs, not just in the effective Hamiltonian model. Our proof is based on the unique topological property of EPs: when the system parameters evolve adiabatically along a loop encircling an EP, the two eigenvalues switch their positions when the system returns to its initial parameters<sup>7,21,25</sup>, in contrast to the typical case where the two eigenvalues return to themselves. Using this property, we numerically

show, in Supplementary Fig. 3 and Supplementary Information section IV, that the complex eigenvalues always switch their positions along every direction in the  $k$  space, and therefore prove the existence of this exceptional ring. As opposed to the simplified effective Hamiltonian model, in a real photonic crystal slab, the EP may exist at a slightly different magnitude of  $k$  and for a slightly different hole radius  $r$  along different directions in the  $k$  space, but this variation is small and negligible in practice (Supplementary Information section V).

To demonstrate the existence of the exceptional ring in such a system, we fabricate large-area periodic patterns in a  $\text{Si}_3\text{N}_4$  slab ( $n = 2.02$  in the visible spectrum, thickness 180 nm) on top of 6  $\mu\text{m}$  of silica ( $n = 1.46$ ) using interference photolithography<sup>24</sup>. Scanning electron microscope (SEM) images of the sample are shown in Fig. 2a, featuring a square lattice (periodicity  $a = 336$  nm) of cylindrical air holes with radius 109 nm. We immerse the structure into an optical liquid with a specified refractive index that can be tuned; accidental degeneracy in the Hermitian part is achieved when the liquid index is selected to be  $n = 1.48$ . We perform angle-resolved reflectivity measurements (set-up shown in Fig. 2b) between  $0^\circ$  and  $2^\circ$  along the  $\Gamma \rightarrow X$  direction and the  $\Gamma \rightarrow M$  direction, for both s and p polarizations. Details of the sample fabrication and the experimental setup can be found in Supplementary Information section VI. The measured reflectivity for the relevant polarization is plotted in the upper panel of Fig. 2c, showing good agreement with numerical simulation results (lower panel), with differences coming from scattering due to surface roughness, inhomogeneous broadening, and the uncertainty in the measurements of system parameters. The complete experimental result for both polarizations is shown in Supplementary Fig. 4; the third and dispersionless band shows up in the other polarization, decoupled from the two bands of interest.

The peaks of reflectivity (dark red colour in Fig. 2c) follow the linear Dirac dispersion; this feature disappears for structures with different radii that do not reach accidental degeneracy (experimental results in



**Figure 2 | Experimental reflectivity spectrum and accidental Dirac dispersion.** **a**, SEM images of the photonic crystal samples: side view (upper panel) and top view (lower panel). **b**, Schematic drawing of the measurement set-up. Linearly polarized light from a super-continuum source is reflected off the photonic crystal slab ('sample') immersed in an optical liquid, and collected by a spectrometer (SP). The incident angle  $\theta$  is controlled using a precision rotating stage. BS, beam splitter. **c**, Reflectivity spectrum of the sample measured experimentally (upper panel) and calculated numerically (lower panel) along the  $\Gamma \rightarrow X$  and the  $\Gamma \rightarrow M$  directions. The peak location of reflectivity reveals the Hermitian part of the system, which forms Dirac dispersion due to accidental degeneracy. In the lower panel, white solid lines

indicate the real part of the eigenvalues; spectra and eigenvalues at three representative angles (marked by dashed lines and circles) are shown in **d**. **d**, Three line cuts of reflectivity  $R$  from simulation results. Also shown are the complex eigenvalues (open circles) calculated numerically. At large angles ( $0.8^\circ$ ), the two resonances are far apart, so the reflectivity peaks (red arrows) are close to the actual positions of the complex eigenvalues. However, at small angles ( $0.3^\circ$ ,  $0.1^\circ$ ), the coupling between resonances cause the resonance peaks (red arrows) to have much greater separations in frequency compared to the complex eigenvalues. The black arrows mark the dips in reflectivity that correspond to the coupled-resonator induced transparency (CRIT, see text for details).

Supplementary Fig. 5). Note that the reflection peaks do not follow the real part of the complex eigenvalues of the Hamiltonian; in fact they follow the eigenvalues of the Hermitian part of the Hamiltonian, even though the Hamiltonian is non-Hermitian. To understand this, we consider a more general two-by-two Hamiltonian of a coupled resonance system  $H$  and separate it into a Hermitian part  $A$  and an anti-Hermitian part  $-iB$  ( $A$  and  $B$  are both Hermitian)

$$H = \underbrace{\begin{pmatrix} \omega_1 & \kappa \\ \kappa & \omega_2 \end{pmatrix}}_A - i \underbrace{\begin{pmatrix} \gamma_1 & \gamma_{12} \\ \gamma_{12} & \gamma_2 \end{pmatrix}}_{iB} \xrightarrow{\text{eigenvalues}} \begin{pmatrix} \omega_+ & 0 \\ 0 & \omega_- \end{pmatrix} \quad (3)$$

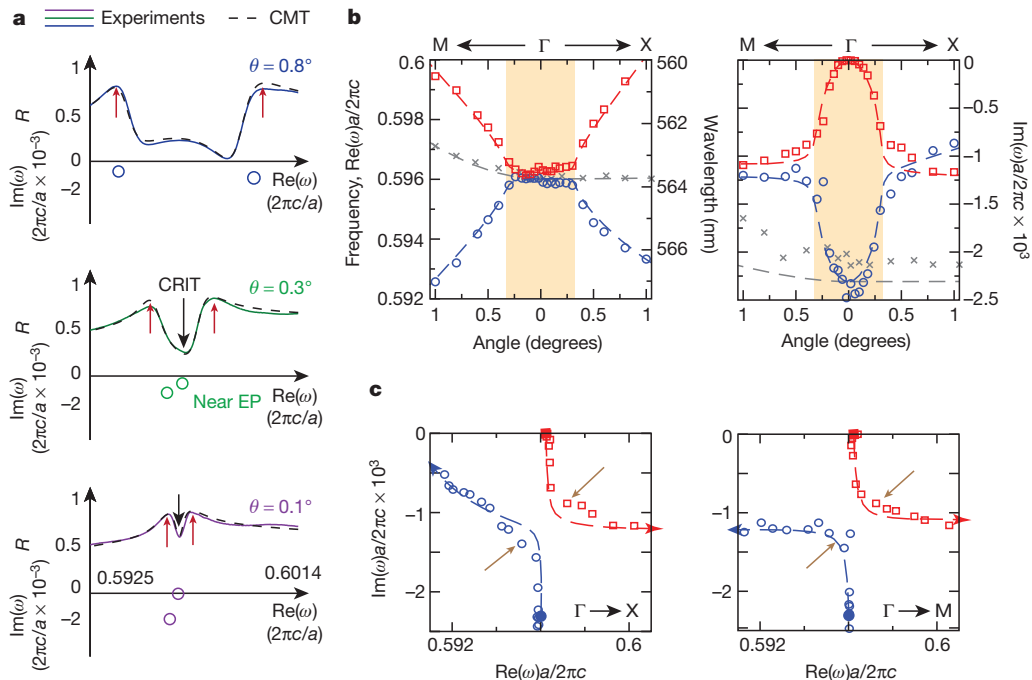
As before, we use  $\omega_{\pm}$  to denote the complex eigenvalues of the Hamiltonian  $A - iB$ . Physically, matrix  $A$  describes a lossless system, and matrix  $-iB$  adds the effects of loss. In  $B$ , the diagonal elements are loss rates (in our system, they come primarily from radiation), and the off-diagonal elements arise from overlap of the two radiation patterns, also known as external coupling of resonances via the continuum. Modelling the reflectivity using temporal coupled-mode theory (TCMT), we show that when matrix  $B$  is dominated by radiation, the reflection peaks occur near the eigenvalues  $\Omega_{1,2}$  of the Hermitian part  $A$  and are independent of the anti-Hermitian part  $-iB$  (see Supplementary Information section VII and Supplementary Fig. 6 for details). Therefore, the linear Dirac dispersion observed in the measured data of Fig. 2c (dark red) indicates that we have successfully achieved accidental degeneracy in the eigenvalues of the Hermitian part, consistent with the simplified model in equation (1). In Supplementary Fig. 8b, we plot the values of  $\Omega_{1,2}$  extracted from the

reflectivity data through a more rigorous data analysis using TCMT (described below); the linear dispersion is indeed observed. We note that when there is substantial non-radiative loss or material gain in the system, the reflection peaks no longer follow the eigenvalues of the Hermitian part (see Supplementary Information section VIII and Supplementary Fig. 7).

The real part of the complex eigenvalues of the Hamiltonian,  $\text{Re}(\omega_{\pm})$ , behave very differently from the reflectivity peaks. Simulation results (solid white lines in the lower panel of Fig. 2c) show  $\text{Re}(\omega_{\pm})$  is dispersionless at small angles with a branch-point singularity around  $0.31^\circ$ —consistent with the feature predicted by the simplified Hamiltonian in equation (2). In Fig. 2d, we compare the reflectivity spectra from simulations (with peaks indicated by red arrows) with the corresponding complex eigenvalues at three representative angles ( $0.8^\circ$  in blue,  $0.31^\circ$  in green and  $0.1^\circ$  in magenta). At  $0.31^\circ$ , the two complex eigenvalues are degenerate, indicating an EP; however, the two reflection peaks do not coincide since they represent the eigenvalues of only the Hermitian part of the Hamiltonian, which does not have degeneracy here. The dip in reflectivity between the two peaks (marked as black arrows in Figs 2 and 3) is the coupled-resonator-induced transparency (CRIT) that arises from the interference between radiation of the two resonances<sup>26</sup>, similar to electromagnetically induced transparency.

Qualitatively, the peak locations of the measured reflectivity spectrum reveal the eigenvalues of the Hermitian part,  $A$ , and the line-widths of the peaks reveal the anti-Hermitian part,  $-iB$ ; diagonalizing  $A - iB$  yields the eigenvalues  $\omega_{\pm}$ , as illustrated in equation (3). To be more quantitative, we use TCMT and account for both the direct





**Figure 3 | Experimental demonstration of an exceptional ring.** **a**, Examples of reflection spectrum  $R$  from the sample at three different angles ( $0.8^\circ$  blue,  $0.3^\circ$  green and  $0.1^\circ$  magenta, solid lines) measured with s-polarized light along the  $\Gamma \rightarrow X$  direction (same setup as in numerical simulations shown in Fig. 2d), fitted with the TCMT expression (equation (S20) in Supplementary Information) (black dashed lines). At each angle, the positions of the complex eigenvalues extracted experimentally are shown as open circles. **b**, Complex eigenvalues extracted experimentally (symbols), with comparison to numerical simulation results (dashed lines) for both the real part (left panel) and the

imaginary part (right panel). Red squares and dashed lines are used for the band with zero radiation loss at the  $\Gamma$  point, blue circles and dashed lines for the band with finite radiation loss at the  $\Gamma$  point, and grey crosses and dashed lines for the third band decoupled from the previous two owing to symmetry. The orange shaded regions correspond to the inside of the ring. **c**, Positions of the eigenvalues (red and blue dashed lines) approach and become very close to each other (indicated by the two brown arrows), demonstrating near-EP features in different directions in the momentum space and the existence of an exceptional ring.

(non-resonant) and the resonant reflection processes including nearby resonances; the expression for reflectivity is given in Supplementary Information equation (S20), with the full derivation given in Supplementary Information section IX. Fitting the reflectivity curves with the TCMT expression gives us an accurate estimate of the matrix elements and the eigenvalues; this procedure is the same as our approach in ref. 27 except that here we additionally account for the coupling between resonances<sup>28</sup>. Figure 3a compares the fitted and the measured reflectivity curves at three representative angles (with more comparison in Supplementary Fig. 8a); the excellent agreement shows the validity of the TCMT model. Underneath the reflectivity curves, we show the complex eigenvalues. The difference between numerically calculated reflectivity (Fig. 2d) and experimental results (Fig. 3a) stems from the non-radiative decay channels in our system, mostly due to scattering loss from the surface roughness<sup>24</sup>.

Repeating the fitting procedure for the reflectivity spectrum measured at different angles, we obtain the dispersion curves for all complex eigenvalues, which are plotted in Fig. 3b. Along both directions in  $k$  space ( $\Gamma \rightarrow X$  and  $\Gamma \rightarrow M$ ), the two bands of interest (shown in blue and red) exhibit the EP behaviour predicted in equation (2): for  $|\mathbf{k}| < k_c$  the real parts are degenerate and dispersionless; for  $|\mathbf{k}| > k_c$  the imaginary parts are degenerate and dispersionless; for  $|\mathbf{k}|$  in the vicinity of  $k_c$  branching features are observed in the real or imaginary part. In Fig. 3c, we plot the eigenvalues on the complex plane for both the  $\Gamma \rightarrow X$  and  $\Gamma \rightarrow M$  directions. We can see that in both directions, the two eigenvalues approach each other and become very close at a certain  $k$  point, which is a clear signature of the system being very close to an EP.

We have shown that non-Hermiticity arising from radiation can significantly alter fundamental properties of the system, including the band structures and the density of states; this effect becomes most

prominent near EPs. The photonic crystal slab described here provides a simple-to-realize platform for studying the influence of EPs on light-matter interaction, such as for single particle detection<sup>21</sup> and modulation of quantum noise. The two-dimensional flat band can also provide a high density of states and therefore high Purcell factors. The strong dispersion of loss in the vicinity of the  $\Gamma$  point can improve the performance of large-area single-mode photonic crystal lasers<sup>29</sup>. The deformation into an exceptional ring is a general phenomenon that can also be achieved with material gain or loss and for Dirac points in other lattices<sup>19,20</sup>. Further studies could advance the understanding of the connection between the topological property of Dirac points<sup>30</sup> and that of EPs<sup>25</sup> in general non-Hermitian wave systems, and our method could go beyond photons to phonons, electrons and atoms.

Received 2 April; accepted 29 June 2015.

Published online 9 September 2015.

- Castro Neto, A. H., Guinea, F., Peres, N. M. R., Novoselov, K. S. & Geim, A. K. The electronic properties of graphene. *Rev. Mod. Phys.* **81**, 109–162 (2009).
- Rechtsman, M. C. *et al.* Strain-induced pseudomagnetic field and photonic Landau levels in dielectric structures. *Nature Photon.* **7**, 153–158 (2013).
- Tarruell, L., Greif, D., Uehlinger, T., Jotzu, G. & Esslinger, T. Creating, moving and merging Dirac points with a Fermi gas in a tunable honeycomb lattice. *Nature* **483**, 302–305 (2012).
- Huang, X., Lai, Y., Hang, Z. H., Zheng, H. & Chan, C. Dirac cones induced by accidental degeneracy in photonic crystals and zero-refractive-index materials. *Nature Mater.* **10**, 582–586 (2011).
- Moitra, P. *et al.* Realization of an all-dielectric zero-index optical metamaterial. *Nature Photon.* **7**, 791–795 (2013).
- Moiseyev, N. *Non-Hermitian Quantum Mechanics* (Cambridge Univ. Press, 2011).
- Rotter, I. A non-Hermitian Hamiltonian operator and the physics of open quantum systems. *J. Phys. A* **42**, 153001 (2009).
- Bender, C. M. & Boettcher, S. Real spectra in non-Hermitian Hamiltonians having PT symmetry. *Phys. Rev. Lett.* **80**, 5243–5246 (1998).
- Rüter, C. E. *et al.* Observation of parity-time symmetry in optics. *Nature Phys.* **6**, 192–195 (2010).

10. Chong, Y., Ge, L. & Stone, A. D. PT-symmetry breaking and laser-absorber modes in optical scattering systems. *Phys. Rev. Lett.* **106**, 093902 (2011).
11. Regensburger, A. *et al.* Parity-time synthetic photonic lattices. *Nature* **488**, 167–171 (2012).
12. Guo, A. *et al.* Observation of PT-symmetry breaking in complex optical potentials. *Phys. Rev. Lett.* **103**, 093902 (2009).
13. Lin, Z. *et al.* Unidirectional invisibility induced by PT-symmetric periodic structures. *Phys. Rev. Lett.* **106**, 213901 (2011).
14. Peng, B. *et al.* Parity-time-symmetric whispering-gallery microcavities. *Nature Phys.* **10**, 394–398 (2014).
15. Liertzer, M. *et al.* Pump-induced exceptional points in lasers. *Phys. Rev. Lett.* **108**, 173901 (2012).
16. Hodaei, H., Miri, M.-A., Heinrich, M., Christodoulides, D. N. & Khajavikhan, M. Parity-time-symmetric microring lasers. *Science* **346**, 975–978 (2014).
17. Feng, L., Wong, Z. J., Ma, R.-M., Wang, Y. & Zhang, X. Single-mode laser by parity-time symmetry breaking. *Science* **346**, 972–975 (2014).
18. Berry, M. Physics of nonhermitian degeneracies. *Czech. J. Phys.* **54**, 1039–1047 (2004).
19. Makris, K., El-Ganainy, R., Christodoulides, D. & Musslimani, Z. H. Beam dynamics in PT symmetric optical lattices. *Phys. Rev. Lett.* **100**, 103904 (2008).
20. Szameit, A., Rechtsman, M. C., Bahat-Treidel, O. & Segev, M. PT-symmetry in honeycomb photonic lattices. *Phys. Rev. A* **84**, 021806 (2011).
21. Cao, H. & Wiersig, J. Dielectric microcavities: model systems for wave chaos and non-Hermitian physics. *Rev. Mod. Phys.* **87**, 61–111 (2015).
22. Sakoda, K. Proof of the universality of mode symmetries in creating photonic Dirac cones. *Opt. Express* **20**, 25181–25194 (2012).
23. Chan, C., Hang, Z. H. & Huang, X. Dirac dispersion in two-dimensional photonic crystals. *Adv. Optoelectron.* **2012**, 313984 (2012).
24. Lee, J. *et al.* Observation and differentiation of unique high-Q optical resonances near zero wave vector in macroscopic photonic crystal slabs. *Phys. Rev. Lett.* **109**, 067401 (2012).
25. Dembowski, C. *et al.* Experimental observation of the topological structure of exceptional points. *Phys. Rev. Lett.* **86**, 787–790 (2001).
26. Hsu, C. W., DeLacy, B. G., Johnson, S. G., Joannopoulos, J. D. & Soljacic, M. Theoretical criteria for scattering dark states in nanostructured particles. *Nano Lett.* **14**, 2783–2788 (2014).
27. Hsu, C. W. *et al.* Observation of trapped light within the radiation continuum. *Nature* **499**, 188–191 (2013).
28. Suh, W., Wang, Z. & Fan, S. Temporal coupled-mode theory and the presence of non-orthogonal modes in lossless multimode cavities. *IEEE J. Quantum Electron.* **40**, 1511–1518 (2004).
29. Chua, S.-L., Lu, L., Bravo-Abad, J., Joannopoulos, J. D. & Soljačić, M. Larger-area single-mode photonic crystal surface-emitting lasers enabled by an accidental Dirac point. *Opt. Lett.* **39**, 2072–2075 (2014).
30. Lu, L., Joannopoulos, J. D. & Soljacic, M. Topological photonics. *Nature Photon.* **8**, 821–829 (2014).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank T. Savas for fabrication of the samples, and F. Wang, Y. Yang, N. Rivera, S. Skirlo, O. Miller and S. G. Johnson for discussions. This work was partly supported by the Army Research Office through the Institute for Soldier Nanotechnologies under contract nos W911NF-07-D0004 and W911NF-13-D-0001. B.Z., L.L. and M.S. were partly supported by S3TEC, an Energy Frontier Research Center funded by the US Department of Energy under grant no. DE-SC0001299. L.L. was supported in part by the Materials Research Science and Engineering Center of the National Science Foundation (award no. DMR-1419807). I.K. was supported in part by Marie Curie grant no. 328853-MC-BSiCS.

**Author Contributions** All authors discussed the results and made critical contributions to the work.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to B.Z. (bozhen@mit.edu) and C.W.H. (chiawei.hsu@yale.edu).

# Inhomogeneity of charge–density–wave order and quenched disorder in a high- $T_c$ superconductor

G. Campi<sup>1,2\*</sup>, A. Bianconi<sup>1,2\*</sup>, N. Poccia<sup>2,3</sup>, G. Bianconi<sup>4</sup>, L. Barba<sup>5</sup>, G. Arrighetti<sup>5</sup>, D. Innocenti<sup>2,6</sup>, J. Karpinski<sup>6,7</sup>, N. D. Zhigadlo<sup>7</sup>, S. M. Kazakov<sup>7,8</sup>, M. Burghammer<sup>9,10</sup>, M. v. Zimmermann<sup>11</sup>, M. Sprung<sup>11</sup> & A. Ricci<sup>2,11\*</sup>

It has recently been established that the high-transition-temperature (high- $T_c$ ) superconducting state coexists with short-range charge-density-wave order<sup>1–11</sup> and quenched disorder<sup>12,13</sup> arising from dopants and strain<sup>14–17</sup>. This complex, multiscale phase separation<sup>18–21</sup> invites the development of theories of high-temperature superconductivity that include complexity<sup>22–25</sup>. The nature of the spatial interplay between charge and dopant order that provides a basis for nanoscale phase separation remains a key open question, because experiments have yet to probe the unknown spatial distribution at both the nanoscale and mesoscale (between atomic and macroscopic scale). Here we report micro X-ray diffraction imaging of the spatial distribution of both short-range charge-density-wave ‘puddles’ (domains with only a few wavelengths) and quenched disorder in  $\text{HgBa}_2\text{CuO}_{4+y}$ , the single-layer cuprate with the highest  $T_c$ , 95 kelvin (refs 26–28). We found that the charge-density-wave puddles, like the steam bubbles in boiling water, have a fat-tailed size distribution that is typical of self-organization near a critical point<sup>19</sup>. However, the quenched disorder, which arises from oxygen interstitials, has a distribution that is contrary to the usually assumed random, uncorrelated distribution<sup>12,13</sup>. The interstitial-oxygen-rich domains are spatially anticorrelated with the charge-density-wave domains, because higher doping does not favour the stripy charge-density-wave puddles, leading to a complex emergent geometry of the spatial landscape for superconductivity.

Although it is known that the incommensurate charge-density-wave (CDW) order in cuprates (copper oxides) is made of ordered, stripy, nanoscale ‘puddles’ with an average of only 3–4 oscillations, information about the size distribution and spatial organization of these puddles has so far not been available. We present experiments that demonstrate that CDW puddles have a complex spatial distribution and coexist with, but are spatially anticorrelated to, quenched disorder in  $\text{HgBa}_2\text{CuO}_{4+y}$  (Hg1201). The sample we studied is a layered perovskite at optimum doping with oxygen interstitials ( $y=0.12$ ), tetragonal symmetry ( $P4/mmm$ ) and a low misfit strain<sup>14–16</sup>. The X-ray diffraction (XRD) measurements (see Methods) show diffuse CDW satellites (secondary peaks surrounding a main peak) at  $\mathbf{q}_{\text{CDW}} = (0.23a^*, 0.16c^*)$  in the  $b^* = 0$  plane and  $\mathbf{q}_{\text{CDW}} = (0.23b^*, 0.16c^*)$  in the  $a^* = 0$  plane (where  $a^*$ ,  $b^*$  and  $c^*$  are the reciprocal lattice units) around specific Bragg peaks, such as (108), below the onset temperature  $T_{\text{CDW}} = 240$  K (see Fig. 1a). The component of the momentum transfer  $\mathbf{q}_{\text{CDW}}$  in the  $\text{CuO}_2$  plane ( $0.23a^*$ ) in this case is smaller than it is in the underdoped case ( $0.28a^*$ )<sup>5</sup>. The temperature evolution of the CDW-peak profile along  $a^*$  (in the  $h$  direction; Fig. 1b) shows a smeared, glassy-like evolution for temperatures below  $T_{\text{CDW}}$ . The CDW-peak intensity reaches a maximum at

$T = 100$  K, followed by a drop associated with the onset of superconductivity at  $T = T_c$ . We investigated the isotropic character of the CDW in the  $a$ – $b$  plane using azimuthal scans, as shown in Fig. 1c. We observed an equal probability of vertically and horizontally striped CDW puddles.

Our main result is the discovery of the statistical spatial distribution of the CDW-puddle size and density throughout the sample, which shows an emergent complex network geometry for the superconducting phase. We performed scanning micro X-ray diffraction ( $\mu\text{XRD}$ ) measurements (see Methods) to extend the imaging of spatial inhomogeneity previously obtained by scanning tunnelling microscopy (STM)<sup>7–9</sup> from the surface to the bulk of the sample and from nanoscale to mesoscale spatial inhomogeneity. Clear evidence of the inhomogeneous spatial distribution of the CDW is provided by the observation of very different CDW-peak profiles collected at different illuminated sample spots (see Fig. 1d) corresponding to spots with ‘large’ and ‘small’ puddles.

We investigated the temperature dependence of CDW domains by recording the CDW-peak intensity and its full-width at half-maximum (FWHM) during cooling from 280 K to 85 K. We collected the data in two different places on the sample corresponding to ‘large’ and ‘small’ CDW puddles. Figure 1e, f shows the temperature dependence of population (intensity), the number of oscillations  $h_{\text{CDW}}/\Delta h_{\text{CDW}}$  (where  $h_{\text{CDW}}$  and  $\Delta h_{\text{CDW}}$  are the position and the FWHM of the CDW peak profile in units of  $a^*$ , respectively) and in-plane puddle size  $\xi_a$  (along the  $a$  axis) in large (red filled circles) and small (black filled squares) CDW puddles. The broad phase transition appears to be arrested, as indicated by the size of the CDW puddles  $\xi_a = 1/\Delta h_{\text{CDW}}$ , which does not diverge below  $T_{\text{CDW}}$ . This behaviour is typical of low-dimensional systems with quenched disorder. A map representing the spatial organization of the CDW-puddle size is shown in Fig. 1g. The probability density function (PDF) of the in-plane CDW-puddle size  $\xi_a$  is shown in Fig. 1h.

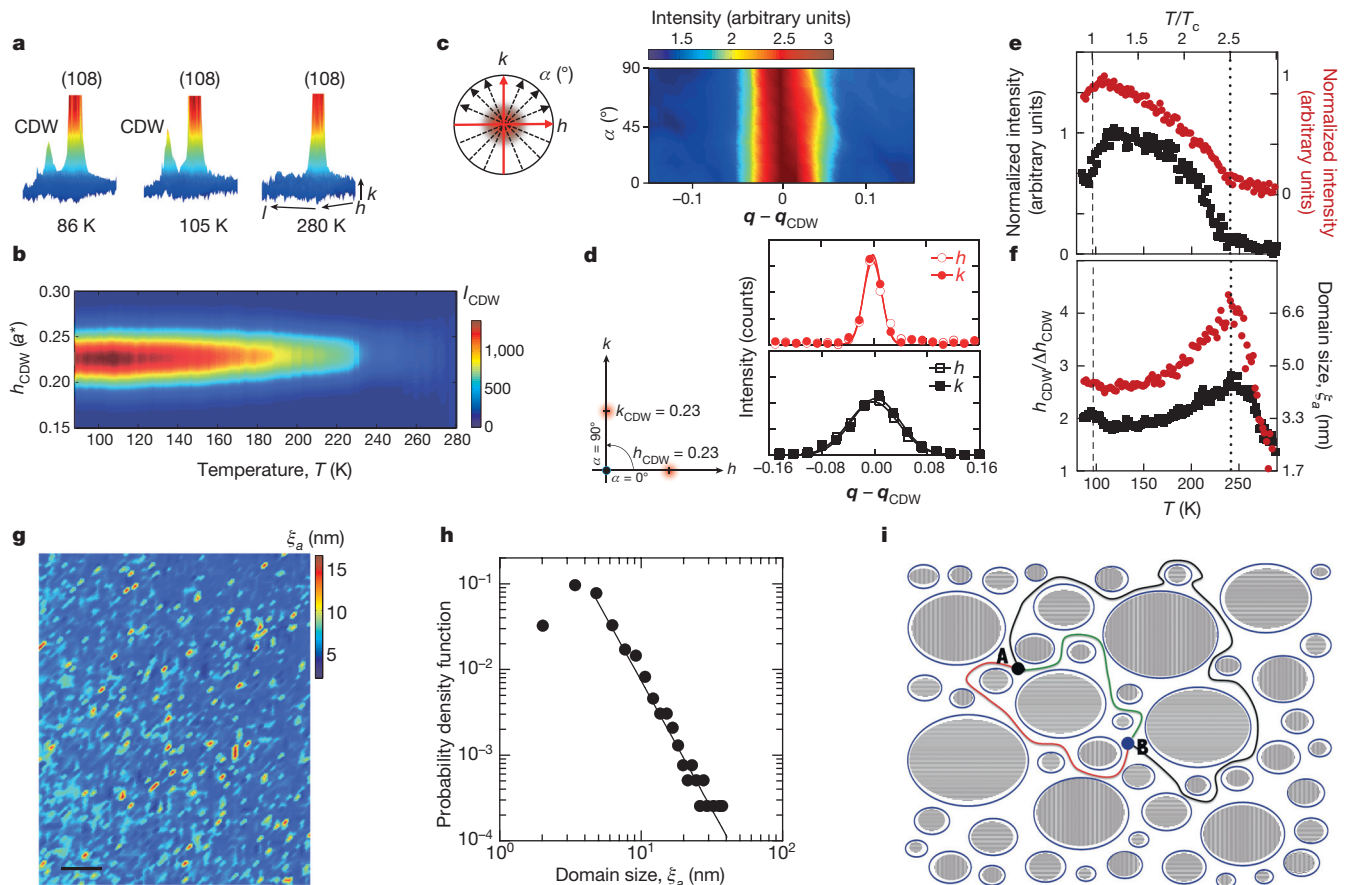
The PDF has a long fat tail that extends over an order of magnitude, and is fitted by  $\text{PDF}(\xi_a) \approx \xi_a^{-\alpha_{\text{CDW}}} \exp(-\xi_a/\xi_t)$ , where  $\alpha_{\text{CDW}} = 2.8 \pm 0.1$  is the critical exponent of the puddle-size power-law distribution and  $\xi_t > 40$  nm. Although we can determine that the average size of CDW puddles is 4.3 nm (in agreement with previous work),  $\text{PDF}(\xi_a)$  has a non-Gaussian shape and rare, larger puddles reaching sizes of 40 nm are detected. Our finding of a fat-tailed distribution for the CDW-puddle size is in agreement with previous results obtained by STM<sup>9</sup>. Such structures, where spontaneous breaking of both translational symmetry (CDW electronic crystalline phase) and gauge symmetry (superconductivity) coexist, have been called superstripes<sup>19</sup>. The distribution of the CDW puddles we have found introduces a substantial topological change to the available space for

<sup>1</sup>Institute of Crystallography, CNR, via Salaria Km 29.300, Monterotondo Roma, I-00015, Italy. <sup>2</sup>Rome International Center for Materials Science, Superstripes, RICMASS, via dei Sabelli 119A, I-00185 Roma, Italy. <sup>3</sup>MESA+ Institute for Nanotechnology, University of Twente, PO Box 217, 7500 AE Enschede, The Netherlands. <sup>4</sup>School of Mathematics, Queen Mary University of London, London E1 4SN, UK.

<sup>5</sup>Institute of Crystallography, Sincrotrone Elettra UOS Trieste, Strada Statale 14 - Km 163,5 Area Science Park, 34149 Basovizza, Trieste, Italy. <sup>6</sup>EPFL, Institute of Condensed Matter Physics, Lausanne CH-1015, Switzerland. <sup>7</sup>ETH, Swiss Federal Institute of Technology Zurich Laboratory for Solid State Physics, CH-8093 Zurich, Switzerland. <sup>8</sup>Department of Chemistry, M.V. Lomonosov Moscow State University, Moscow 119991, Russia. <sup>9</sup>European Synchrotron Radiation Facility, BP 220, F-38043 Grenoble Cedex, France. <sup>10</sup>Department of Analytical Chemistry, Ghent University, Krijgslaan 281, S12 B-9000 Ghent, Belgium. <sup>11</sup>Deutsches Elektronen-Synchrotron DESY, Notkestraße 85, D-22607 Hamburg, Germany.

\*These authors contributed equally to this work.





**Figure 1 | Temperature dependence and spatial distribution of CDW puddles in Hg1201.** **a**, The CDW satellite near the (108) Bragg peak appears below 240 K. **b**, Temperature dependence of CDW-peak profiles along  $h$ . The CDW-peak intensity  $I_{\text{CDW}}$  is measured as the number of counts minus the background. **c**, The  $q_{\text{CDW}} = (0.23a^*(b^*), 0.16c^*)$  peak profile at different azimuthal angles  $\alpha$  showing the peak isotropy. **d**, Two typical CDW peaks collected at two different places in the same crystal. Red solid circles correspond to the diffraction profile from an illuminated part the sample with large CDW puddles (red in **g**); black filled squares correspond to an illuminated part of the sample with small CDW puddles (blue in **g**); and the solid lines are

Gaussian fits. **e**, The CDW-peak intensity as a function of temperature, at the two different places on the sample corresponding to large (red filled circles, right axis) and small (black filled squares, left axis) CDW puddles. The dashed line corresponds to  $T = T_c$  and the dotted line to  $T = T_{\text{CDW}}$ . **f**, Evolution of the number of CDW oscillations ( $h_{\text{CDW}}/\Delta h_{\text{CDW}}$ ) inside a CDW puddle and the CDW domain size along the  $a$  axis ( $\xi_a$ ). **g**, **h**, Spatial map (**g**) and probability density function (**h**) of the CDW-puddle size. Scale bar in **g**, 10  $\mu\text{m}$ . **i**, A schematic of non-equivalent paths, running in the interface space between CDW puddles, connecting point A to point B in the emergent complex non-Euclidean spatial geometry<sup>29,30</sup> for the superconducting current.

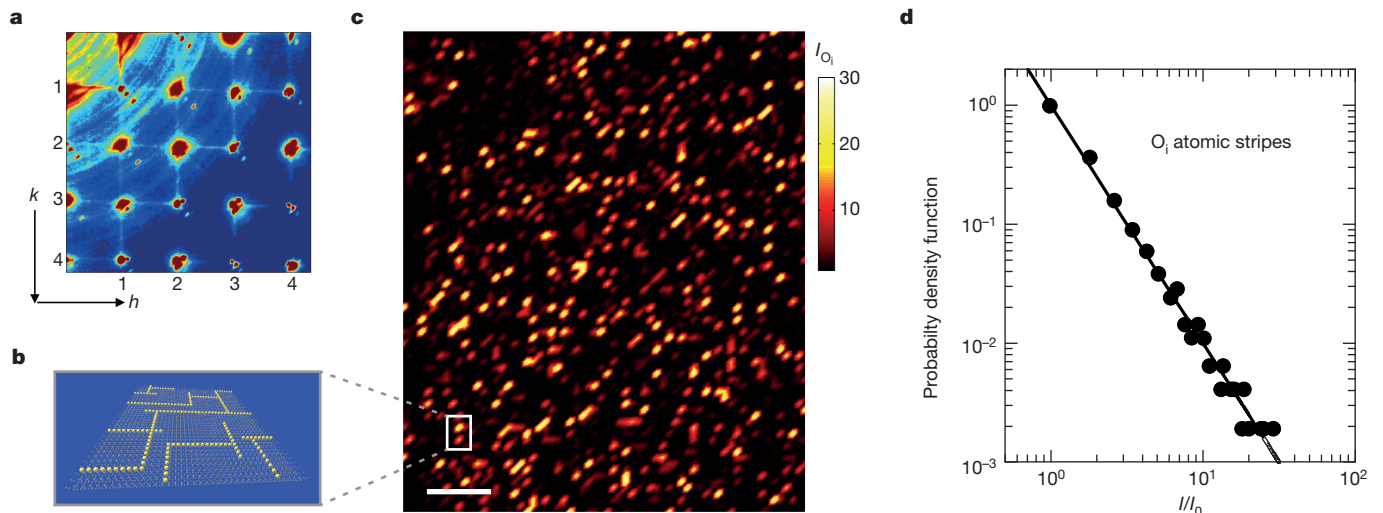
superconductivity: the current running from a point A to a point B of the material can take different paths (see Fig. 1i) that are not topologically equivalent<sup>29</sup> thus forming an emergent complex hyperbolic geometry<sup>30</sup>.

To investigate the interplay between the CDW puddles and the quenched disorder, we studied the spatial distribution of oxygen defects. The quenched lattice disorder is due to oxygen interstitials ( $\text{O}_i$ ), which form  $\text{O}_i$  atomic stripes in the  $\text{HgO}_y$  layers, in agreement with previous experiments<sup>27,28</sup>.  $\text{HgBa}_2\text{CuO}_{4+y}$  (ref. 28), like  $\text{YBa}_2\text{Cu}_3\text{O}_{6+y}$  (ref. 15) and  $\text{La}_2\text{CuO}_{4+y}$  (refs 14, 16), shows  $T_c$  variations, owing to the effect of the spatial organization of  $\text{O}_i$  on superconductivity. The average  $\text{O}_i$  self-organization was detected by high-energy XRD (see Methods). Figure 2a shows the ( $0 < h < 5$ ,  $0 < k < 4$ ) portion of reciprocal space, where there is strong evidence of diffuse streaks running along the  $a^*$  and  $b^*$  directions and crossing all the Bragg peaks. Our high-energy XRD data confirm the formation of  $\text{O}_i$  stripes intercalated between the  $\text{CuO}_2$  planes, both in the (100) and (010) directions<sup>26</sup>.

The spatial distribution of the intensity of the streaks was obtained by  $\text{S}\mu\text{XRD}$  (see Methods). We measured the reciprocal  $a^*-c^*$  plane (or  $b^*-c^*$  plane) around the (006) Bragg peak in reflection geometry. The  $\text{O}_i$  stripes in Hg1201 run along the  $a^*$  ( $b^*$ ) direction with no correlation along the  $c^*$  direction; therefore, they also lead to streaks

on the  $a^*-c^*$  plane. A schematic of  $\text{O}_i$  atomic stripes is shown in Fig. 2b. In Fig. 2c we show the spatial map of the streak intensity. The picture shows rich (bright yellow) and poor (dark black) regions of  $\text{O}_i$  stripes. The PDF of  $\text{O}_i$ -rich regions in Fig. 2d is fitted by  $\text{PDF}(I) \approx (I/I_0)^{-\alpha_{\text{O}_i}} \exp(I/I_t)$ , where  $I_0$  is the average intensity,  $\alpha_{\text{O}_i} = 2.0 \pm 0.1$  is the critical exponent and  $I_t > 20$ .

In Fig. 3 we present results on the spatial interplay between CDW-rich regions and  $\text{O}_i$ -rich regions. We calculated the ‘difference map’ (see Methods) between CDW peaks and  $\text{O}_i$  diffuse streaks. The poor CDW regions on the  $\text{CuO}_2$  basal plane correspond to  $\text{O}_i$ -rich regions on the  $\text{HgO}_y$  layers, as illustrated in Fig. 3a. The CDW puddles and  $\text{O}_i$ -rich regions give rise to the positive and negative peaks, respectively, in the surface plot shown in Fig. 3b. The spatial anticorrelation is evident from the scatter plot of  $\text{O}_i$  intensity versus CDW intensity (Fig. 3c). As  $\text{O}_i$  intensity increases, the CDW intensity decreases, and vice versa. This is consistent with the fact that excess  $\text{O}_i$  means higher doping, and high doping does not favour stripy, underdoped short-range CDW order. Figure 3d shows the two maps obtained via the segmentation of the difference map, and provides a direct image of how doping-poor (CDW-rich regions are shown in red) and doping-rich ( $\text{O}_i$ -rich regions are shown in blue) phases are arranged in different regions of the material. Figure 3e illustrates the nanoscale configuration of CDW-puddles (red spots) in the  $\text{CuO}_2$  plane using

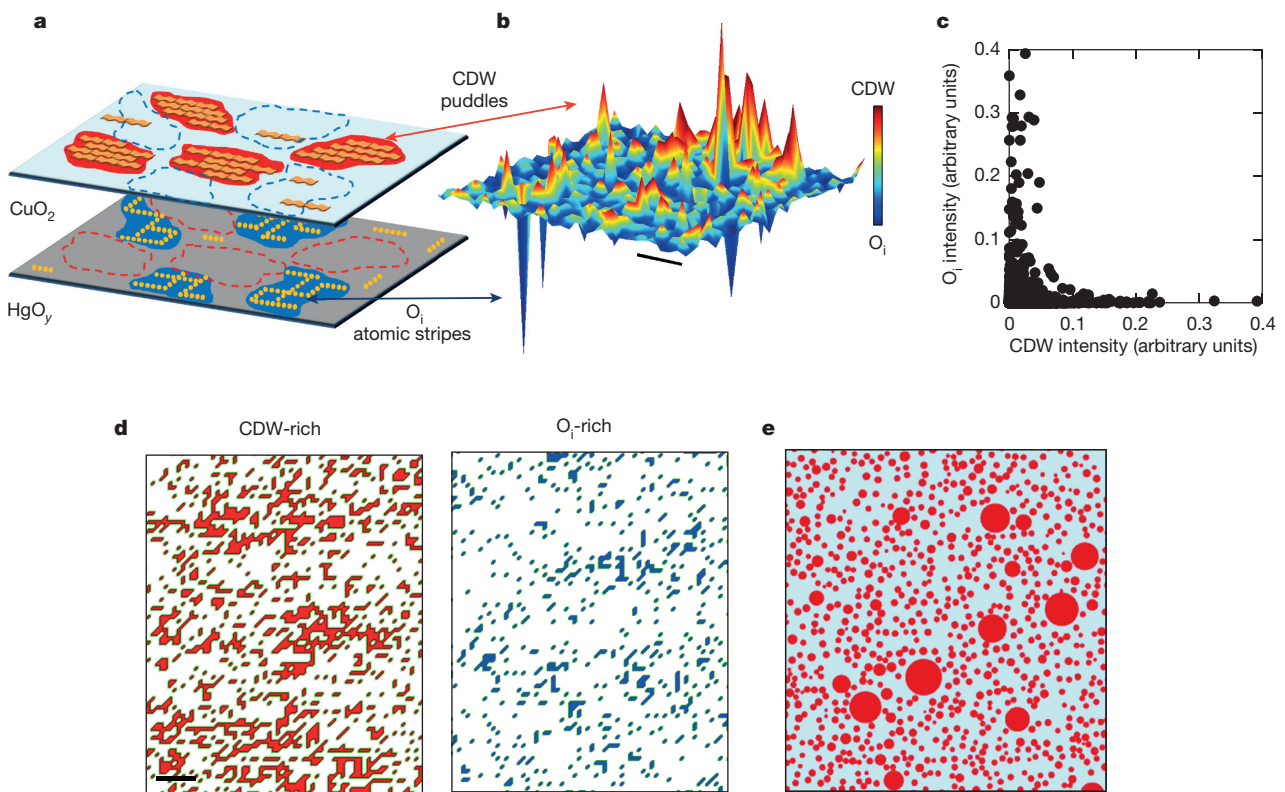


**Figure 2 | Correlated quenched disorder due to  $O_i$  atomic stripes in Hg1201.** **a**, A portion of the  $h$ - $k$  diffraction pattern. Resolution-limited streaks connect the Bragg peaks, owing to the formation of  $O_i$  stripes in the  $HgO_y$  spacer layers. **b**, Schematic representation of the atomic  $O_i$  stripes. **c**,  $S\mu$ XRD

map of a region of **a** showing the relative  $O_i$  streak intensity  $I_{O_i}$ . The bright (dark) spots correspond to sample regions with a high (low) density of  $O_i$  atomic stripes, called  $O_i$ -rich (poor) regions. Scale bar, 10  $\mu$ m. **d**, Probability density function calculated from the  $O_i$ -streaks intensity map.

the experimental distribution of CDW size; this distribution generates 'holes' in the space available for the free electrons (light blue area). This space is topologically interesting: there are an infinite number of ways for a current path to connect a point A to a point B around the CDW puddles, which are not only distinguished by the number of times a path goes around a single hole, but also by the way the path passes

though the pattern of CDW puddles<sup>27,28</sup>. The complex space that emerges from the mesoscopic phase separation, both in the spacer layers and in the  $CuO_2$  plane, substantially changes (1) the dielectric constant that controls the long-range Coulomb interaction that is relevant for phase separation near a Lifshitz transition<sup>19</sup>, (2) the dielectric constant that is relevant to electron-electron interaction in the



**Figure 3 | Spatial anticorrelation between CDW-rich and  $O_i$ -rich regions.** **a**, The CDW-rich regions (red) on the  $CuO_2$  planes and  $O_i$ -rich regions (blue) on the  $HgO_y$  layers. **b**, Surface plot of the difference map (see Methods) between the CDW-peak and  $O_i$ -streak intensity. The positive (green to red) values indicate the CDW-rich regions and the negative (green to blue) values correspond to  $O_i$ -rich regions. Scale bar, 5  $\mu$ m. **c**, Scatter plot of  $O_i$  versus CDW intensity demonstrating the negative correlation between CDW-puddle and

$O_i$ -stripe populations. **d**, Segmentations of the difference map in **b** highlighting the network of CDW-rich domains (left panel) and  $O_i$ -rich regions (right panel). Scale bar, 10  $\mu$ m. **e**, A schematic of the nanoscale texture formed by CDW-rich regions (red spots) and the 'charge- $O_i$ -rich' region (light blue area), which define an interface space and loci of the superconductivity with a complex non-Euclidean geometry<sup>29,30</sup>.

pairing and (3) the geometrical and topological properties of the space that is available for the overall phase coherence of the macroscopic quantum condensate that is made up of multiple condensates at the nanoscale with a single critical temperature<sup>19,23</sup>.

This work offers new insight into the complexity of nanoscale phase-separation phenomena in high-temperature superconductors. More generally, our results deal with the effects of quenched disorder in phase transitions. A phase transition that would be first order in the clean limit gets smeared into a continuous-looking transition in the presence of a random, Gaussian distributed, quenched disordered background<sup>12,13</sup>. Here the disorder itself is not randomly distributed, but has a long-tailed probability density function, leading to correlated disorder. Even in the ‘ideal’ single-layer cuprate superconductor  $\text{HgBa}_2\text{CuO}_{4+y}$  at optimum doping ( $T_c = 95\text{ K}$ ), the CDW order self-organizes into puddles, forming an inhomogeneous landscape with an emergent complex network geometry. Our results provide further evidence for the universality of mesoscale phase separation even in the most optimized superconducting cuprates, which implies that the superconductivity will be non-uniform throughout what is a granular medium.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 22 March; accepted 21 July 2015.**

- Chang, J. *et al.* Direct observation of competition between superconductivity and charge density wave order in  $\text{YBa}_2\text{Cu}_3\text{O}_{6.67}$ . *Nature Phys.* **8**, 871–876 (2012).
- Croft, T. P., Lester, C., Senn, M. S., Bombardi, A. & Hayden, S. M. Charge density wave fluctuations in  $\text{La}_{2-x}\text{Sr}_x\text{CuO}_4$  and their competition with superconductivity. *Phys. Rev. B* **89**, 224513 (2014).
- Comin, R. *et al.* Broken translational and rotational symmetry via charge stripe order in underdoped  $\text{YBa}_2\text{Cu}_3\text{O}_{6+y}$ . *Science* **347**, 1335–1339 (2015).
- Poccia, N. *et al.* Optimum inhomogeneity of local lattice distortions in  $\text{La}_2\text{CuO}_{4+y}$ . *Proc. Natl Acad. Sci. USA* **109**, 15685–15690 (2012).
- Tabis, W. *et al.* Charge order and its connection with Fermi-liquid charge transport in a pristine high- $T_c$  cuprate. *Nature Commun.* **5**, 5875 (2014).
- Comin, R. *et al.* Charge order driven by Fermi-arc instability in  $\text{Bi}_2\text{Sr}_{2-x}\text{La}_x\text{CuO}_{6+\delta}$ . *Science* **343**, 390–392 (2014).
- Kohsaka, Y. *et al.* An intrinsic bond-centered electronic glass with unidirectional domains in underdoped cuprates. *Science* **315**, 1380–1385 (2007).
- Mesaros, A. *et al.* Topological defects coupling smectic modulations to intra-unit-cell nematicity in cuprates. *Science* **333**, 426–430 (2011).
- Phillabaum, B., Carlson, E. W. & Dahmen, K. A. Spatial complexity due to bulk electronic nematicity in a superconducting underdoped cuprate. *Nature Commun.* **3**, 915 (2012).
- Gabovich, A. M., Voitenko, A. I., Annett, J. F. & Ausloos, M. Charge- and spin-density-wave superconductors. *Supercond. Sci. Technol.* **14**, R1–R27 (2001).
- Bianconi, A. *et al.* Determination of the local lattice distortions in the  $\text{CuO}_2$  plane of  $\text{La}_{1.85}\text{Sr}_{0.15}\text{CuO}_4$ . *Phys. Rev. Lett.* **76**, 3412–3415 (1996).
- Imry, Y. & Ma, S.-K. Random-field instability of the ordered state of continuous symmetry. *Phys. Rev. Lett.* **35**, 1399–1401 (1975).
- Vojta, T. Rare region effects at classical, quantum and nonequilibrium phase transitions. *J. Phys. Math. Gen.* **39**, R143–R205 (2006).
- Frattini, M. *et al.* Scale-free structural organization of oxygen interstitials in  $\text{La}_2\text{CuO}_{4+y}$ . *Nature* **466**, 841–844 (2010).
- Ricci, A. *et al.* Multiscale distribution of oxygen puddles in 1/8 doped  $\text{YBa}_2\text{Cu}_3\text{O}_{6.67}$ . *Sci. Rep.* **3**, 2383 (2013).
- Poccia, N. *et al.* Evolution and control of oxygen order in a cuprate superconductor. *Nature Mater.* **10**, 733–736 (2011).
- Drees, Y. *et al.* Hour-glass magnetic excitations induced by nanoscopic phase separation in cobalt oxides. *Nature Commun.* **5**, 5731 (2014).
- Gor'kov, L. P. & Teitel'baum, G. B. Two-component energy spectrum of cuprates in the pseudogap phase and its evolution with temperature and at charge ordering. *Sci. Rep.* **5**, 8524 (2015).
- Bianconi, A. Shape resonances in superstripes. *Nature Phys.* **9**, 536–537 (2013).
- Alvarez, G., Moreo, A. & Dagotto, E. Complexity in high-temperature superconductors. *Low Temp. Phys.* **32**, 290–297 (2006).
- Kresin, V., Ovchinnikov, Y. & Wolf, S. Inhomogeneous superconductivity and the “pseudogap” state of novel superconductors. *Phys. Rep.* **431**, 231–259 (2006); erratum **437**, 233–234 (2007).
- She, J.-H. & Zaanen, J. BCS superconductivity in quantum critical metals. *Phys. Rev. B* **80**, 184518 (2009).
- Bianconi, G. Superconductor-insulator transition on annealed complex networks. *Phys. Rev. B* **85**, 061113 (2012).
- Davison, R. A., Schalm, K. & Zaanen, J. Holographic duality and the resistivity of strange metals. *Phys. Rev. B* **89**, 245116 (2014).
- Lucas, A. & Sachdev, S. Conductivity of weakly disordered strange metals: from conformal to hyperscaling-violating regimes. *Nucl. Phys. B* **892**, 239–268 (2015).
- Karpinski, J. *et al.* High-pressure synthesis, crystal growth, phase diagrams, structural and magnetic properties of  $\text{Y}_2\text{Ba}_4\text{Cu}_7\text{O}_{2n+x}$ ,  $\text{HgBa}_2\text{Ca}_{n-1}\text{Cu}_n\text{O}_{2n+2+\delta}$  and quasi-one-dimensional cuprates. *Supercond. Sci. Technol.* **12**, R153–R181 (1999).
- Wagner, J. L. *et al.* in *Phase Transitions and Self-Organization in Electronic and Molecular Networks* (eds Thorpe, M. F. & Phillips, J. C.) 331–339 (Springer, 2001).
- Izquierdo, M. *et al.* One dimensional ordering of doping oxygen in  $\text{HgBa}_2\text{CuO}_{4+\delta}$  superconductors evidenced by X-ray diffuse scattering. *J. Phys. Chem. Solids* **72**, 545–548 (2011).
- Zeng, W., Sarkar, R., Luo, F., Gu, X. & Gao, J. Resilient routing for sensor networks using hyperbolic embedding of universal covering space. In *INFOCOM, 2010 Proc. IEEE 1–9* (IEEE, 2010).
- Wu, Z., Menichetti, G., Rahmede, C. & Bianconi, G. Emergent complex network geometry. *Sci. Rep.* **5**, 10073 (2015).

**Acknowledgements** We acknowledge the ESRF, ELETTRA and DESY synchrotron facilities for radiation-beam time and support. We thank the beamline scientists for help with experiments. We acknowledge the Calypso programme for travel support. We acknowledge support from the Superstripes Institute. N.P. acknowledges financial support from a Marie Curie Intra-European Fellowship for career development.

**Author Contributions** All authors have contributed to essential portions of this work. The experiment was conceived by A.B., G.C., N.P., G.B. and A.R.; the  $\text{Hg1201}$  crystals were grown at ETH by S.M.K. and J.K.; N.D.Z. and N.P. performed magnetic characterization of  $\text{Hg1201}$  single crystals; experiments at DESY were performed by A.R., G.C., N.P., M.S., M.v.Z. and D.I.; experiments at ESRF were performed by A.R., N.P., G.C., A.B. and M.B.; experiments at Elettra were performed by G.C., L.B., G.A., A.B. and A.R.; and the data analysis was carried out by G.C., A.B., G.B. and A.R. All authors discussed the results and contributed to the writing of the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.B. ([antonio.bianconi@ricmass.eu](mailto:antonio.bianconi@ricmass.eu)).



## METHODS

**Sample preparation and characterization.** The  $\text{HgBa}_2\text{CuO}_{4+y}$  (Hg1201) crystal with  $y = 0.12$ , grown at ETH<sup>24</sup>, has a sharp superconducting transition at  $T_c = 95$  K. The crystal structure has  $P4/mmm$  symmetry with lattice parameters  $a = b = 0.387480(5)$  nm and  $c = 0.95078(2)$  nm at  $T = 100$  K (numbers in parentheses indicate the standard deviation of the last digit).

**XRD measurements using the XRD1 beamline.** To identify the CDW order in a single Hg1201 crystal we used XRD using the XRD1 beamline at the Elettra synchrotron radiation facility in Trieste, Italy, tuning the photon energy between 13 keV and 16 keV with a beam size of  $200 \times 200 \mu\text{m}^2$ . Only selected reflections show clear CDW satellites, in agreement with ref. 2. We focused on the CDW satellite located at  $q_{\text{CDW}} = (0.23, 0, 0.16)$  around the (108) Bragg reflection, which appeared as the sample was cooled below 240 K. Typical diffraction patterns collected at 85 K, 105 K and 280 K are shown in Fig. 1a. To get a direct view of the temperature dependence of the CDW-satellite reflection for  $T = 280$ –85 K, a two-dimensional colour plot of the CDW-peak profile along the  $a^*$  direction as a function of temperature is shown in Fig. 1b.

**High-energy XRD measurements using the BW5 beamline.** High-energy XRD measurements were collected using the BW5 beamline at DESY, Hamburg, Germany, using a transmission geometry and an X-ray energy of 100 KeV. A single SiGe(111) gradient monochromator was used. The beam size was  $200 \mu\text{m} \times 200 \mu\text{m}$ . We used a vertical rotation axis, and the  $c$  axis of the single crystal was oriented parallel to the direction of the incoming X-ray beam. The diffraction patterns were collected by an area detector in the temperature range 20–300 K. In this geometry, we can probe the lattice fluctuations on the  $a$ – $b$  plane. The possible CDW-peak anisotropy was seen using azimuthal scans ( $0^\circ < \alpha < 90^\circ$ ), as shown in the schematic in Fig. 1c. The CDW-peak amplitude and FWHM do not change substantially as a function of  $\alpha$ , as shown in the colour plot of the diffraction profile (as a function of  $\alpha$ ) in Fig. 1c; instead, this plot shows the CDW planar isotropy, in agreement with the tetragonal  $P4/mmm$  symmetry of the lattice. The presence of resolution-limited streaks connecting the Bragg peaks, owing to the organization of single  $\text{O}_i$  stripes in the mercury spacer layer, are shown in Fig. 2a. This figure shows a portion of the  $h$ – $k$  diffraction pattern that was collected at DESY. The spatial distribution of the  $\text{O}_i$  atomic stripes in the  $\text{HgO}_y$  spacer layers, which cause one-dimensional doping and lattice spatial inhomogeneity, does not vary with temperature below 250 K, leading to quenched disorder at the onset of the charge-density phase.

**S $\mu$ XRD measurements using the ID13 beamline.** S $\mu$ XRD experiments were performed in reflection geometry using the ID13 beamline at ESRF, Grenoble, France. We applied incident X-ray energy of 13 KeV. By moving the sample under a 1- $\mu\text{m}$  focused beam with an  $x$ – $y$  translator, we scanned a sample area of  $65 \times 80 \mu\text{m}^2$ , collecting 5,200 different diffraction patterns at  $T = 100$  K. For each scanned point of the sample, the  $q_{\text{CDW}}$ -peak profile was extracted; the FWHMs along the  $a^*(b^*)$  and  $c^*$  directions were evaluated to obtain the domain size of the charge-ordered regions along the  $a(b)$  and  $c$  crystallographic axes.

Two quite different CDW-peak profiles in the same crystal measured along the  $a^*(b^*)$  direction in the  $a^*-c^*(b^*-c^*)$  plane are shown in Fig. 1d. Here we show two typical profiles collected at two different spatial locations in the same crystal corresponding to large (red circles) and small (black squares) puddles. The continuous lines are the Gaussian fits to the data. The different amplitudes and FWHMs ( $(0.033 \pm 0.001)a^*$  and  $(0.089 \pm 0.001)a^*$  in the upper and lower panels of Fig. 1d, respectively, errors indicate standard deviation) of the two peaks, which correspond to large and small CDW puddles, provide evidence of a strong inhomogeneity in the CDW spatial distribution. The peak profiles do appear the same along the  $a^*$  and  $b^*$  directions, confirming the peak isotropy in the basal plane of the tetragonal lattice. The intensity of the CDW satellites as a function of temperature, measured at two different locations on the sample corresponding to large (red) and small (black) CDW puddles, are shown in Fig. 1e. The vertical lines represent the superconducting temperature  $T_c$  and the CDW onset temperature  $T_{\text{CDW}}$ . The order–disorder transition is very broad, which indicates the role of the quenched disorder owing to the presence of defects. Moreover, the CDW intensity shows a clear drop around  $T_c$  that appears to depend on the CDW puddle size. The temperature dependence of the number of CDW oscillations inside a single puddle ( $h_{\text{CDW}}/\Delta h_{\text{CDW}}$ ) and the domain size of a single puddle along the  $a(b)$  axis ( $\xi_a$ ) are shown in Fig. 1f. ( $\Delta h_{\text{CDW}}$  and  $h_{\text{CDW}}$  are the FWHM and the location along  $a^*$  of the CDW peak; the domain size along the  $a$  axis ( $b$  axis) is given by the correlation length  $\xi_a$ .) The inhomogeneity of the CDW distribution is depicted in the  $65 \times 80 \mu\text{m}^2$  XRD map of the (nanoscale) size of CDW domains in Fig. 1g. This figure shows loci of large (red–yellow area) and small (blue area) CDW puddles. The scale bar corresponds to 10  $\mu\text{m}$ . Using the ID13 microfocus beamline at ESRF, we can also detect the spatial distribution of the quenched disorder. Figure 2c shows the S $\mu$ XRD map of the integrated intensity of the streaks of  $\text{O}_i$  stripes. The bright (dark) spots correspond to sample regions with a high (low) density of  $\text{O}_i$  atomic stripes, called  $\text{O}_i$ -rich (poor) regions. The scale bar is 10  $\mu\text{m}$ . Figure 2d shows the PDF of the  $\text{O}_i$ -streak intensity that was obtained from the S $\mu$ XRD map. This plot shows the probability distribution of the  $\text{O}_i$ -rich regions. The experimental set-up allows us to investigate the spatial interplay between CDW puddles in the  $\text{CuO}_2$  plane and  $\text{O}_i$ -rich domains in the  $\text{HgO}_y$  layers, shown in Fig. 3. We measured the ‘difference map’  $\langle I_{\text{CDW}} - I_{\text{O}_i} \rangle$ , where  $\langle I_{\text{CDW}} \rangle$  and  $\langle I_{\text{O}_i} \rangle$  are the intensities of the  $q_{\text{CDW}}$  peak and the  $\text{O}_i$  diffuse streaks, respectively, normalized to [0, 1]. The surface plot of this difference map is shown in Fig. 3b. The positive (green to red) peaks indicate CDW-puddle-rich regions and the negative (green to blue) peaks indicate  $\text{O}_i$ -rich regions. The spatial anticorrelation between CDW puddles and  $\text{O}_i$  atomic stripes is obtained by segmentation of the difference map. We use this segmentation to visualize the phase separation owing to the network of CDW-rich domains, which correspond to ‘charge poor’ domains in the  $\text{CuO}_2$  planes (left panel of Fig. 3d), and  $\text{O}_i$ -rich regions in the  $\text{HgO}_y$  layers, which correspond to ‘charge rich’ portions of the  $\text{CuO}_2$  plane (right panel of Fig. 3d).

**Code availability.** The code we used for statistical analysis of the S $\mu$ XRD data is not currently available (G.C., A.R. and A.B., manuscript in preparation).

# Designing switchable polarization and magnetization at room temperature in an oxide

P. Mandal<sup>1</sup>, M. J. Pitcher<sup>1</sup>, J. Alaria<sup>2</sup>, H. Niu<sup>1</sup>, P. Borisov<sup>1†</sup>, P. Stamenov<sup>3</sup>, J. B. Claridge<sup>1</sup> & M. J. Rosseinsky<sup>1</sup>

Ferroelectric and ferromagnetic materials exhibit long-range order of atomic-scale electric or magnetic dipoles that can be switched by applying an appropriate electric or magnetic field, respectively. Both switching phenomena form the basis of non-volatile random access memory<sup>1</sup>, but in the ferroelectric case, this involves destructive electrical reading and in the magnetic case, a high writing energy is required<sup>2</sup>. In principle, low-power and high-density information storage that combines fast electrical writing and magnetic reading can be realized with magnetoelectric multiferroic materials<sup>3</sup>. These materials not only simultaneously display ferroelectricity and ferromagnetism, but also enable magnetic moments to be induced by an external electric field, or electric polarization by a magnetic field<sup>4,5</sup>. However, synthesizing bulk materials with both long-range orders at room temperature in a single crystalline structure is challenging because conventional ferroelectricity requires closed-shell  $d^0$  or  $s^2$  cations, whereas ferromagnetic order requires open-shell  $d^n$  configurations with unpaired electrons<sup>6</sup>. These opposing requirements pose considerable difficulties for atomic-scale design strategies such as magnetic ion substitution into ferroelectrics<sup>7,8</sup>. One material that exhibits both ferroelectric and magnetic order is BiFeO<sub>3</sub>, but its cycloidal magnetic structure<sup>9</sup> precludes bulk magnetization and linear magnetoelectric coupling<sup>10</sup>. A solid solution of a ferroelectric and a spin-glass perovskite combines switchable polarization<sup>11</sup> with glassy magnetization, although it lacks long-range magnetic order<sup>12</sup>. Crystal engineering of a layered perovskite has recently resulted in room-temperature polar ferromagnets<sup>13</sup>, but the electrical polarization has not been switchable. Here we combine ferroelectricity and ferromagnetism at room temperature in a bulk perovskite oxide, by constructing a percolating network of magnetic ions with strong superexchange interactions within a structural scaffold exhibiting polar lattice symmetries at a morphotropic phase boundary<sup>14</sup> (the compositional boundary between two polar phases with different polarization directions, exemplified by the PbZrO<sub>3</sub>–PbTiO<sub>3</sub> system) that both enhances polarization switching and permits canting of the ordered magnetic moments. We expect this strategy to allow the generation of a range of tunable multiferroic materials.

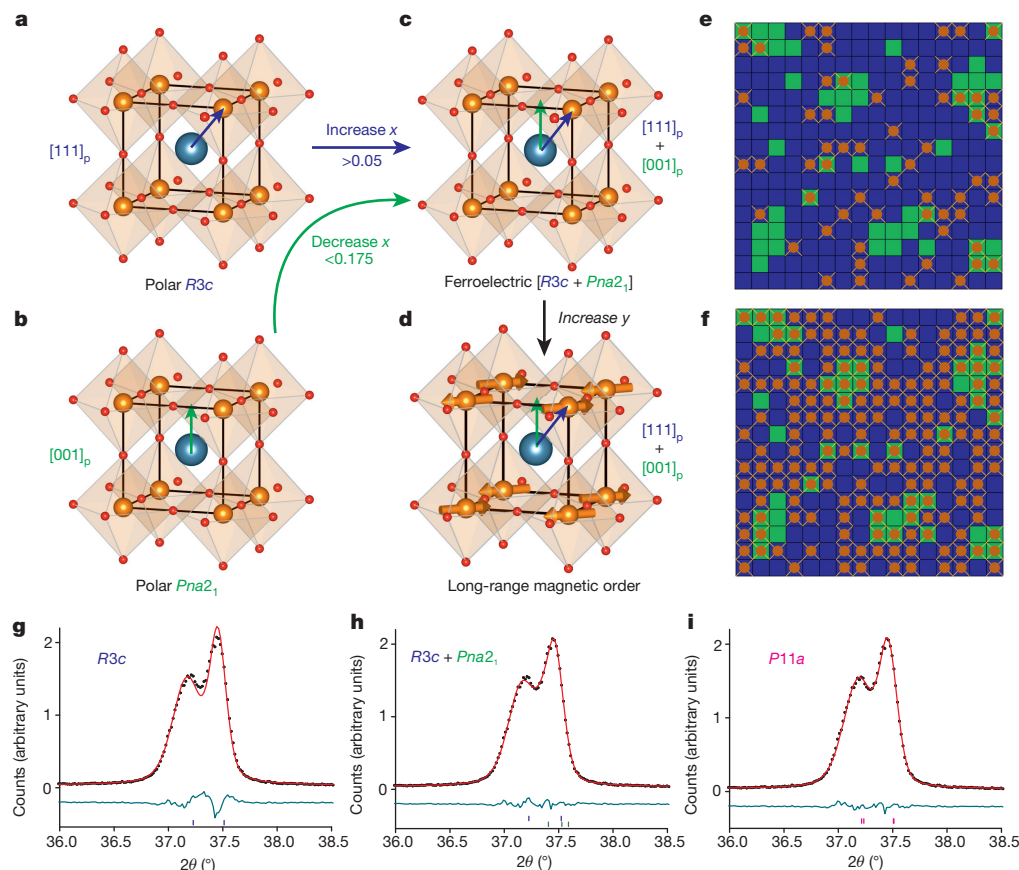
Several approaches to room-temperature multiferroicity have been explored. Composite multiferroics, which are multiphase mixtures of magnetic and ferroelectric materials, have displayed the largest magnetoelectric effects, originating from stress-mediated coupling<sup>15</sup>. The indirect nature of the cross-coupling between the polar and magnetic phases hinders complete switching of the ferroic properties through magnetoelectric coupling. The single-phase oxide BiFeO<sub>3</sub> is antiferromagnetically ordered with competing exchange interactions producing a cycloidal structure with a period of 62 nm (ref. 9). Two approaches have been used to disrupt this cycloid. First, solid solutions of the non-polar, weakly ferromagnetic LnFeO<sub>3</sub> (Ln = Sm, Dy, La) ferrites in BiFeO<sub>3</sub> have a finite magnetization at room temperature<sup>16,17</sup> in a fully ordered magnetic network. The inherent trade-off between the soft magnetic properties of the orthoferrite and

the ferroelectric properties of BiFeO<sub>3</sub> leads to intermediate compositions for which the long-range crystallographic symmetry (polar versus non-polar)<sup>18</sup>, the magnetic ground state and switchability<sup>19</sup> are subject to debate. Second, strained and nanostructured BiFeO<sub>3</sub> films have shown remanent magnetization<sup>20</sup>, and electrical control of the staggered magnetization in BiFeO<sub>3</sub> can switch the magnetization of a coupled ferromagnetic material in a thin-film-device structure<sup>21</sup>.

In BiFeO<sub>3</sub>, the ferroelectric polarization is aligned along the [111]<sub>p</sub> direction of the primitive cubic (indicated by the subscript p) ABO<sub>3</sub> perovskite subcell. The morphotropic phase boundary (MPB) between two non-cubic, polar crystallographic symmetries of the ABO<sub>3</sub> perovskite with distinct polarization directions is a route to large, switchable polarization via polarization rotation or reorientation<sup>14</sup>. The structure at the MPB is a single perovskite network with a complex domain microstructure<sup>22</sup>, where the Bragg scattering can be modelled in single- or multiple-phase approximations<sup>23</sup>. We have recently produced a new MPB in a solid solution between rhombohedral (R, space group *R3c*, Fig. 1a) [111]<sub>p</sub> and orthorhombic (O, space group *Pna2*<sub>1</sub>, Fig. 1b) [001]<sub>p</sub> polarization directions in the Bi<sup>3+</sup>-based perovskites  $(1-x)\text{BiTi}_{(1-y)/2}\text{Fe}_y\text{Mg}_{(1-y)/2}\text{O}_3-(x)\text{CaTiO}_3$ ; the MPB occurs for  $0.075 \leq x < 0.175$ ,  $y = 0.25$  (Fig. 1c)<sup>24</sup>. This new MPB affords large switchable polarizations (*P*) in bulk materials, for example,  $P = 49 \mu\text{C cm}^{-2}$  for  $x = 0.15$ ,  $y = 0.25$ . These materials were designed to have high  $d^0$  Ti<sup>4+</sup> and Mg<sup>2+</sup> cation content on the octahedral B site to minimize dielectric loss and aid ferroelectric switching by sustaining the required electric field. Because the MPB structure is based on a continuous ABO<sub>3</sub> network, there is a coherent magnetic B-site sublattice that is connected by B–O–B superexchange pathways throughout each crystallite. The low Fe content of 21.25% at  $x = 0.15$ ,  $y = 0.25$  is below the percolation threshold for the primitive cubic lattice (Fig. 1e): because magnetic order in insulators arises from nearest-neighbour superexchange, such a material cannot display long-range magnetic order, and none is observed (demonstrated by the monotonic temperature dependence of the field-cooled (FC) and zero-field-cooled (ZFC) magnetizations and the linear magnetization (*M*(*H*), where *H* is the magnetic field) isotherm at 100 K, Extended Data Fig. 1).

We therefore explored increasing the Fe content  $((1-x) \times y)$  on the B site to generate long-range magnetic order within an MPB system that displays switchable polarization (Fig. 1d, f). A series of compositions in the range  $x = 0.15$ ,  $0.60 \leq y \leq 0.90$  were prepared and the perovskite phase purity was confirmed by powder X-ray diffraction (PXRD; Extended Data Fig. 2). The compositions  $x = 0.15$ ,  $y = 0.60$  and  $x = 0.15$ ,  $y = 0.80$  were selected for detailed property studies. Pawley refinements on these compositions show that a model with both R and O phases, and a single-phase monoclinic model in a  $\sqrt{2}a_p \times 2a_p \times \sqrt{2}a_p$  unit cell (space group *P11a*, which is a polar subgroup of both *R3c* and *Pna2*<sub>1</sub>; refined lattice parameters shown in Extended Data Table 1), produce superior fits to those obtained by a purely rhombohedral model (Extended Data Fig. 3 and Extended Data Table 1). This result demonstrates that these compositions exist

<sup>1</sup>Department of Chemistry, University of Liverpool, Liverpool L69 7ZD, UK. <sup>2</sup>Department of Physics, University of Liverpool, Liverpool L69 7ZE, UK. <sup>3</sup>CRANN, Trinity College Dublin, College Green, Dublin 2, Ireland. <sup>†</sup>Present address: Department of Physics, West Virginia University, Morgantown, West Virginia 26506, USA.



**Figure 1 | Crystal structure, magnetic percolation and the morphotropic phase boundary (MPB) in  $(1-x)\text{BiTi}_{(1-y)/2}\text{Fe}_y\text{Mg}_{(1-y)/2}\text{O}_3-(x)\text{CaTiO}_3$  where  $0 \leq x \leq 0.35$  and  $0.25 \leq y \leq 0.90$ .** **a**, Schematic diagram of the purely rhombohedral ( $R3c$ ) structure where  $x \leq 0.05$ ,  $y = 0.25$ , represented in the cubic perovskite subcell with polar displacement of Bi along the  $[111]_p$  axis. The blue, orange and red spheres indicate Bi/Ca, Fe/Ti/Mg and O respectively. **b**, The purely orthorhombic ( $Pna2_1$ ) structure where  $0.175 \leq x \leq 0.35$ ,  $y = 0.25$  with polar displacement of Bi along the  $[001]_p$  axis. **c**, The ferroelectric MPB observed for  $0.075 \leq x < 0.175$ ,  $y = 0.25$ , shown with superimposed polar displacements of Bi along the  $[111]_p$  and  $[001]_p$  axes. **d**, Long-range magnetic order at 300 K for  $x = 0.15$ ,  $y = 0.80$ , with a proposed magnetic structure (orange arrows) based on a G-type antiferromagnetic arrangement and spins oriented perpendicular to the  $[111]_p$  polarization direction. **e**, **f**, Schematic

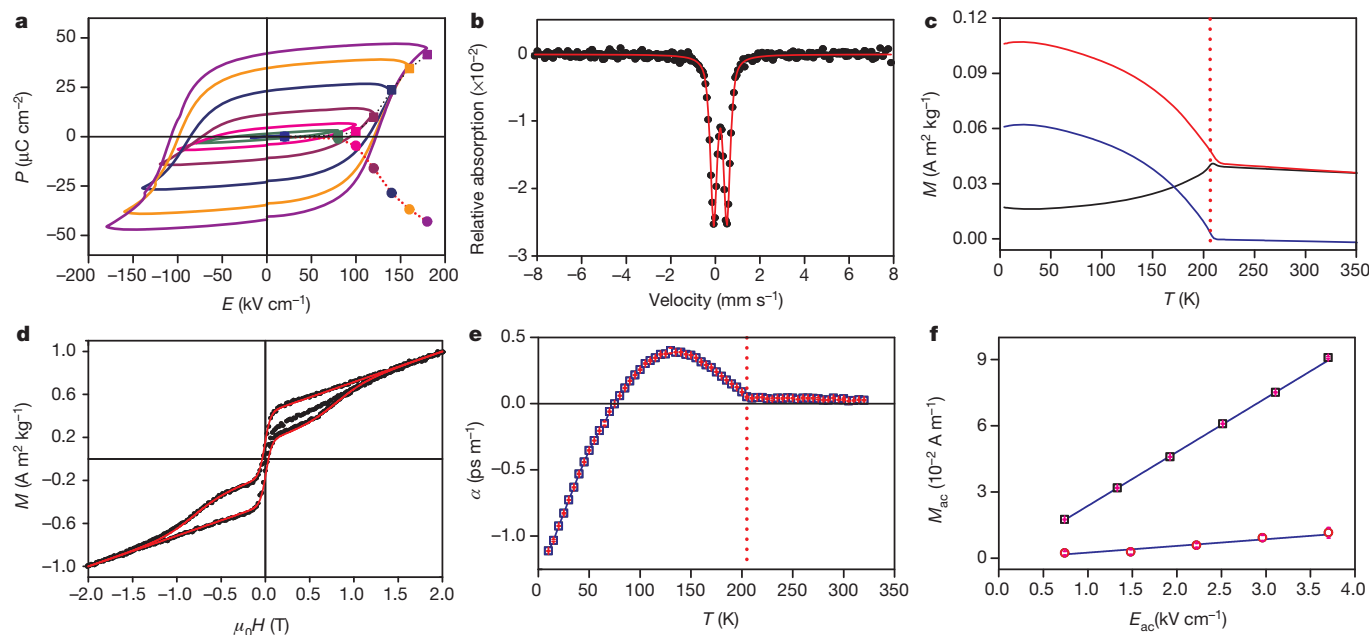
in the MPB region towards the rhombohedral limit, and hence are long-range ordered, polar, non-cubic materials.

The  $x = 0.15$ ,  $y = 0.60$  material has 51% Fe present on the B site, which is above the percolation threshold for long-range magnetic order, and has low dielectric loss despite the enhanced  $d$ -electron content (Extended Data Fig. 4a). It retains the polarization switching characteristics of the MPB and is a ferroelectric at room temperature with a maximum polarization ( $P_{\text{max}}$ ) of  $47.1 \mu\text{C cm}^{-2}$  (Fig. 2a). Positive-up negative-down (PUND) measurements confirm the intrinsic nature of the measured polarization, with a remanent polarization of  $41.5 \mu\text{C cm}^{-2}$ . The 300 K Mössbauer spectrum, which probes all the Fe nuclei in the sample, is a sharp paramagnetic doublet (Fig. 2b), showing that the material is not magnetically ordered at room temperature. The good single paramagnetic component fit (hyperfine field  $B_{\text{hf}} = 0$ ) to the data excludes any magnetically ordered impurity phases with concentrations higher than 2 wt%. The isomer shift of  $\delta = 0.22(3) \text{ mm s}^{-1}$  (where the number in parentheses represents the standard error) corresponds to  $\text{Fe}^{3+}$  in a homogeneous, distorted octahedral environment. The loop observed in the  $M(H)$  isotherm at 300 K, which has a small coercive field, is therefore associated with trace amounts (below diffraction detection limits) of Fe-rich ferrimagnetic impurities (for example,  $\text{Fe}_3\text{O}_4$  or  $\text{MgFe}_2\text{O}_4$  spinels;

diagrams of the MPB microstructure and nearest-neighbour magnetic exchange pathways for  $x = 0.15$ ,  $y = 0.25$  (**e**) and for  $x = 0.15$ ,  $y = 0.80$  (**f**). Each square represents a perovskite unit cell (rhombohedral in purple and orthorhombic in green), brown dots are distributed randomly to represent unit cells containing Fe, and the associated brown lines represent magnetic exchange pathways. A percolating exchange pathway spanning the sample is absent in **e** but present in **f**, **g**–**i**. Pawley fits (red lines) to PXRD data (black circles) in the angular range  $36.0^\circ \leq 2\theta \leq 38.5^\circ$  from composition  $x = 0.15$ ,  $y = 0.80$ , modelled using a single  $R3c$  unit cell (**g**), superimposed  $R3c$  and  $Pna2_1$  unit cells (**h**) and a single monoclinic  $P11a$  unit cell (**i**); see Extended Data Fig. 3 for full patterns. Teal line, difference between the measured and fitted data; purple markers,  $hkl$  ( $R3c$ ) reflections; green markers,  $hkl$  ( $Pna2_1$ ) reflections; magenta markers,  $hkl$  ( $P11a$ ) reflections.

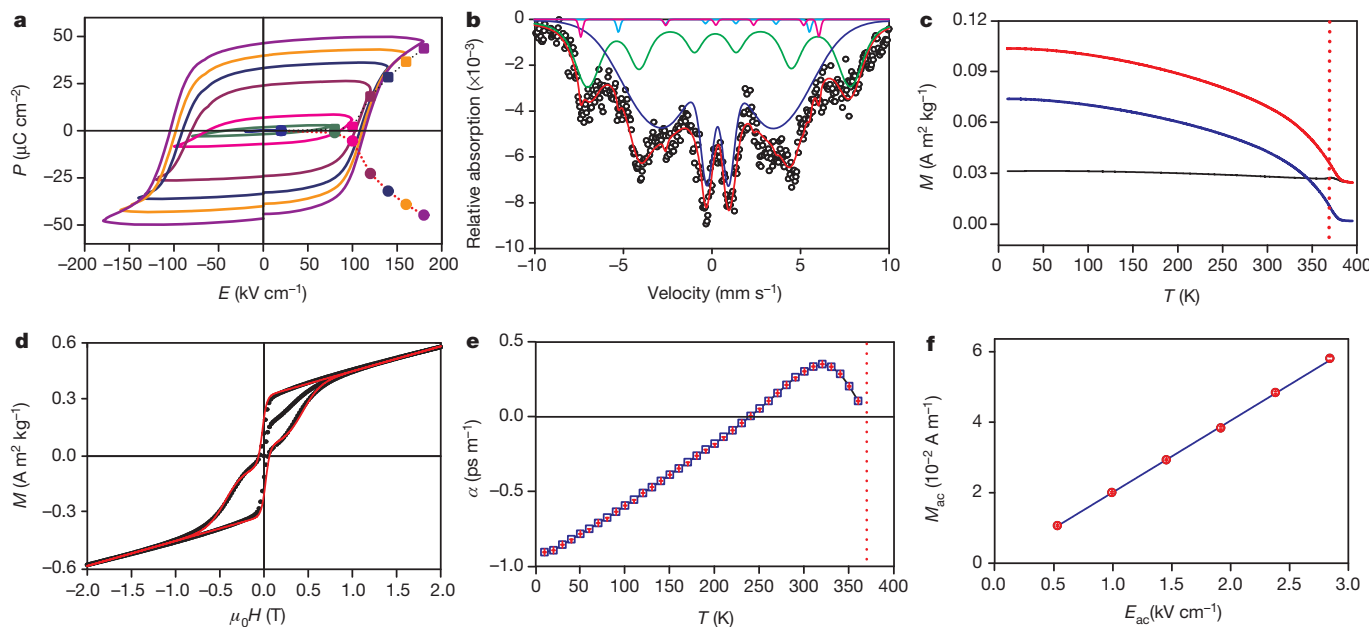
Extended Data Fig. 5), and does not correspond to long-range order of the perovskite. Magnetic ordering in the perovskite below 350 K was probed by dc-SQUID (superconducting quantum interference device) ZFC and FC magnetization, and thermal remanent magnetization (TRM) measurements (Fig. 2c). The large divergence between the ZFC and FC data indicates the onset of weak ferromagnetism at the Néel temperature,  $T_N = 205 \text{ K}$ , consistent with the Brillouin-like drop in the TRM. No other sign of magnetic ordering at lower temperature is observed (Extended Data Figure 6a), suggesting that, at the MPB, the perovskite behaves as a single-phase magnetic material; this suggestion is consistent with the sharpness of the Mössbauer spectrum. The  $M(H)$  isotherm at 10 K (Fig. 2d) can be decomposed into two components, a soft magnetic phase, which is associated with a trace amount (approximately 0.6 wt%), consistent with the 300 K measurement shown in Extended Data Fig. 5c) of the Fe-rich spinel ferrite impurity, and a harder phase with an open hysteresis loop and a linear high-field contribution, which are characteristic features of a weak ferromagnet (Extended Data Fig. 5a). This harder magnetic phase is attributed to the perovskite compound with an extracted coercive field of 376 mT and saturation magnetization of  $0.013 \mu_B$  per Fe, confirming that the material is a weak ferromagnet, where the magnetization arises from ferromagnetic canting of a predominantly antiferromagnetic magnetic





**Figure 2 | Ferroelectric, magnetic and magnetoelectric properties of composition  $x = 0.15$ ,  $y = 0.60$ .** **a**, Polarization  $P$  versus applied electric field  $E$  at 300 K showing ferroelectric switching, measured at 10 Hz. Filled squares and circles represent the remanent polarizations from PUND measurements. Each colour represents the maximum applied electric field in the  $P(E)$ /PUND measurements; dotted lines are included as visual aids. **b**, Mössbauer spectrum measured at 300 K, with no applied magnetic field (black circles) and a single paramagnetic component fit (red line). **c**, dc magnetization measurement of TRM (blue line) and ZFC/FC (black/red lines) magnetization. The red dotted line indicates  $T = T_N$ . The FC and ZFC data only converge at the highest temperature,

because the small divergence above the Néel temperature arises from an impurity phase with an ordering temperature above the highest measured  $T$ . **d**, Isothermal magnetization  $M(H)$  at 10 K (black circles) and the sum of the perovskite and spinel impurity phase contributions (red line). **e**, Temperature dependence of the linear magnetoelectric susceptibility ( $\alpha$ ). The data points (blue squares) are the mean values from 10 repeated measurements, with standard errors shown in red. The red dotted line is  $T = T_N$ . **f**, Induced ac magnetization ( $M_{ac}$ ) versus applied ac electric field amplitude ( $E_{ac}$ ) at 150 K (black squares) and 300 K (red circles). The data points are the mean values from 10 repeated measurements, with standard errors shown in red; the blue lines are linear fits to the data.



**Figure 3 | Ferroelectric, magnetic and magnetoelectric properties of compositions  $x = 0.15$ ,  $y = 0.80$ .** **a**, Polarization  $P$  versus applied electric field  $E$  at 300 K showing ferroelectric switching, measured at 10 Hz. Filled squares and circles represent the remanent polarizations from PUND measurements. Each colour represents the maximum applied electric field in the  $P(E)$ /PUND measurements; dotted lines are included as visual aids. **b**, Mössbauer spectrum measured at 300 K, with no applied magnetic field (black circles) and the multicomponent fit (red line). Individual components 1–4 (green, blue, cyan and magenta, respectively) are described in the text and summarized in Extended Data Table 2. **c**, dc magnetization measurement of TRM (blue line)

and ZFC/FC (black/red lines) magnetization. The red dotted line indicates  $T = T_N$ . **d**, Isothermal magnetization  $M(H)$  at 300 K (black circles) and the sum of the perovskite and minority phase spinel contributions (red line). **e**, Temperature dependence of the linear magnetoelectric susceptibility ( $\alpha$ ). The data points (blue squares) are the mean values from 10 repeated measurements, with standard errors shown in red. The red dotted line is  $T = T_N$ . **f**, Induced ac magnetization ( $M_{ac}$ ) versus applied ac electric field amplitude ( $E_{ac}$ ) at 300 K (red circles). The data points are the mean values from 10 repeated measurements, with standard errors shown in red; the blue line is a linear fit to the data.

structure. The structural symmetries present at the MPB all permit canting to occur within the G-type antiferromagnetic arrangement that is generally found for perovskite ferrites<sup>25</sup> (Fig. 1d).

To confirm whether the two order parameters  $P$  and  $M$  are coupled, magnetoelectric measurements were performed on a disk that was poled both electrically and magnetically. The  $x = 0.15$ ,  $y = 0.60$  material displays linear magnetoelectric coupling (measured as the slope of the induced ac magnetization ( $M_{ac}$ ) versus the applied ac electric field amplitude ( $E_{ac}$ )) only below the long-range ordering temperature of 205 K (Fig. 2e, f). At 10 K, the material shows a pronounced magnetoelectric susceptibility  $\alpha (= \mu_0 M_{ac}/E_{ac})$ , where  $\mu_0$  is the vacuum permeability) of  $-1.11(1) \text{ ps m}^{-1}$  (Extended Data Fig. 7; the number in parentheses represents the standard error), which changes sign upon warming to  $T_N$  (ref. 26). The residual 300 K magnetoelectric coupling is an order of magnitude smaller than that in the magnetically ordered state (Fig. 2f compares data below  $T_N$  at 150 K and above  $T_N$  at 300 K) and can be associated with composite effects<sup>8,15</sup> involving the magnetic minority phases that are not integrated into the complex MPB microstructure of the  $\text{ABO}_3$  perovskite network.

The 51% Fe content at  $x = 0.15$ ,  $y = 0.60$  is sufficient to percolate and give long-range magnetic order, but the mean exchange field is too weak for room-temperature magnetization, because the effective number of nearest neighbours for superexchange is too low. The  $x = 0.15$ ,  $y = 0.80$  composition gives 68% B-site occupancy by  $\text{Fe}^{3+}$ . Mössbauer spectroscopy demonstrates that this increased coverage produces bulk magnetic order at 300 K (Fig. 3b), in contrast to  $x = 0.15$ ,  $y = 0.60$ , because there are no paramagnetic contributions to the spectrum (Extended Data Table 2). ZFC/FC magnetization and TRM measurements (Fig. 3c) show that  $T_N$  increases to 370 K. The majority (98.6(2)%) components 1 and 2 arise from the magnetically ordered MPB perovskite. Component 1 corresponds to  $\text{Fe}^{3+}$  in a slightly distorted octahedral environment ( $\delta = 0.29(5) \text{ mm s}^{-1}$ , electric quadrupole moment  $Q = 0.033(7) \text{ mm s}^{-1}$ ). The broader local field distribution and reduced hyperfine field in component 2 reflect the different local magnetic environments in a percolating system<sup>27</sup>. The minority (1.3(2)%) components arise from spinel-derived  $\text{Fe}^{3+}$  ( $\delta = 0.3 \text{ mm s}^{-1}$ ). There is no signature of further magnetic ordering at lower temperature in the TRM plot (Extended Data Fig. 6b). The 300 K (below  $T_N$ )  $M(H)$  isotherm for the  $x = 0.15$ ,  $y = 0.80$  composition is similar to that observed for  $x = 0.15$ ,  $y = 0.60$  in the magnetically ordered state at 10 K: there are two components, a soft phase that is attributed to the high Fe content impurity (approximately 0.7 wt%), and a harder phase that is assigned to the perovskite with a coercive field (367 mT) and remanent magnetization ( $0.008 \mu_B$  per Fe) consistent with bulk weak ferromagnetic behaviour (Fig. 3d and Extended Data Fig. 6b). The  $x = 0.15$ ,  $y = 0.80$  material is a ferroelectric at room temperature with a switchable maximum polarization ( $P_{\text{max}}$ ) of  $49.9 \mu\text{C cm}^{-2}$  (Fig. 3a): a ferroelectric polarization is still measurable at 473 K (Extended Data Fig. 8). The remanent polarization obtained from PUND measurement is  $43.7 \mu\text{C cm}^{-2}$  (Fig. 3a and Extended Data Fig. 4c). PUND and leakage current measurements confirm the intrinsic origin of the polarization, consistent with the 300 K dc resistivity of  $2.1 \times 10^{12} \Omega \text{ cm}$  (Extended Data Fig. 4d). The switching of the intrinsic perovskite weak ferromagnetic magnetization at the bulk coercive field thus coexists with the switching of the ferroelectric polarization at room temperature.

The long-range ordered  $P$  and  $M$  in  $x = 0.15$ ,  $y = 0.80$  afford bulk magnetoelectric coupling at room temperature with a linear magnetoelectric susceptibility ( $\alpha$ ) of  $0.26(1) \text{ ps m}^{-1}$  (Fig. 3f). Variable temperature measurements (Fig. 3e) show that  $\alpha$  is  $-0.91(1) \text{ ps m}^{-1}$  at 10 K, with a change of sign similar to that found for  $x = 0.15$ ,  $y = 0.60$  upon heating. The linear magnetoelectric susceptibility tends to zero at the bulk  $T_N$  (Fig. 3e), demonstrating that it arises from interaction between the coexisting magnetic and electric long-range orders. The  $x = 0.15$ ,  $y = 0.80$  material with 68%  $\text{Fe}^{3+}$  B-site occupancy (and

thus a percolating network of Fe–O–Fe superexchange paths to give long-range magnetic order) is a room-temperature magnetoelectric ferromagnetic ferroelectric material. The introduction of high-temperature long-range magnetic order into MPB systems is a diversifiable strategy for the generation of tunable multiferroic materials.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 12 March; accepted 29 June 2015.

1. Zhu, J.-G. Magnetoresistive random access memory: the path to competitiveness and scalability. *Proc. IEEE* **96**, 1786–1798 (2008).
2. Bibes, M. Nanoferronics is a winning combination. *Nature Mater.* **11**, 354–357 (2012).
3. Scott, J. F. Data storage: multiferroic memories. *Nature Mater.* **6**, 256–257 (2007).
4. Weisheit, M. *et al.* Electric field-induced modification of magnetism in thin-film ferromagnets. *Science* **315**, 349–351 (2007).
5. Chen, X., Hochstrat, A., Borisov, P. & Kleemann, W. Magnetoelectric exchange bias systems in spintronics. *Appl. Phys. Lett.* **89**, 202508 (2006).
6. Hill, N. A. Why are there so few magnetic ferroelectrics? *J. Phys. Chem. B* **104**, 6694–6709 (2000).
7. Smolenskii, G. A. & Chupis, I. E. Ferroelectromagnets. *Sov. Phys. Usp.* **25**, 475–493 (1982).
8. Fiebig, M. Revival of the magnetoelectric effect. *J. Phys. D* **38**, R123–R152 (2005).
9. Sosnowska, I., Neumaier, T. P. & Steichele, E. Spiral magnetic ordering in bismuth ferrite. *J. Phys. C* **15**, 4835–4846 (1982).
10. Popov, Y. F. *et al.* Linear magnetoelectric effect and phase transitions in bismuth ferrite,  $\text{BiFeO}_3$ . *JETP Lett.* **57**, 69–73 (1993).
11. Evans, D. M. *et al.* Magnetic switching of ferroelectric domains at room temperature in multiferroic PZTFT. *Nature Commun.* **4**, 1534 (2013).
12. Chhail, S. *et al.* Magnetic short- and long-range order in  $\text{PbFe}_{0.5}\text{Ta}_{0.5}\text{O}_3$ . *Phys. Rev. B* **89**, 174418 (2014).
13. Pitcher, M. J. *et al.* Tilt engineering of spontaneous polarization and magnetization above 300 K in a bulk layered perovskite. *Science* **347**, 420–424 (2015).
14. Damjanovic, D. A morphotropic phase boundary system based on polarization rotation and polarization extension. *Appl. Phys. Lett.* **97**, 062906 (2010).
15. Nan, C.-W., Bichurin, M. I., Dong, S., Viehland, D. & Srinivasan, G. Multiferroic magnetoelectric composites: historical perspective, status, and future directions. *J. Appl. Phys.* **103**, 031101 (2008).
16. Yuan, G. L. & Or, S. W. Multiferroicity in polarized single-phase  $\text{Bi}_{0.875}\text{Sm}_{0.125}\text{FeO}_3$  ceramics. *J. Appl. Phys.* **100**, 024109 (2006).
17. Zhang, S. *et al.* Observation of room temperature saturated ferroelectric polarization in Dy substituted  $\text{BiFeO}_3$  ceramics. *J. Appl. Phys.* **111**, 074105 (2012).
18. Arnold, D. Composition-driven structural phase transitions in rare-earth-doped  $\text{BiFeO}_3$  ceramics: a review. *IEEE Trans. Ultrason. Ferr.* **62**, 62–82 (2015).
19. Khomchenko, V. A. *et al.* Structural, ferroelectric and magnetic properties of  $\text{Bi}_{0.85}\text{Sm}_{0.15}\text{FeO}_3$  perovskite. *Cryst. Res. Technol.* **46**, 238–242 (2011).
20. Heron, J. T. *et al.* Electric-field-induced magnetization reversal in a ferromagnet-multiferroic heterostructure. *Phys. Rev. Lett.* **107**, 217202 (2011).
21. Heron, J. T. *et al.* Deterministic switching of ferromagnetism at room temperature using an electric field. *Nature* **516**, 370–373 (2014).
22. Zhang, N. *et al.* The missing boundary in the phase diagram of  $\text{PbZr}_{1-x}\text{Ti}_x\text{O}_3$ . *Nature Commun.* **5**, 5231 (2014).
23. Noheda, B. & Cox, D. E. Bridging phases at the morphotropic boundaries of lead oxide solid solutions. *Phase Transit.* **79**, 5–20 (2006).
24. Mandal, P. *et al.* Morphotropic phase boundary in the Pb-free  $(1-x)\text{Bi}_{1-x/3}\text{Fe}_{2/3}\text{Mg}_{3/8}\text{O}_3-x\text{CaTiO}_3$  system: tetragonal polarization and enhanced electromechanical properties. *Adv. Mater.* **27**, 2883–2889 (2015).
25. White, R. L. Review of recent work on the magnetic and spectroscopic properties of the rare-earth orthoferrites. *J. Appl. Phys.* **40**, 1061–1069 (1969).
26. Shtrikman, S. & Treves, D. Observation of the magnetoelectric effect in  $\text{Cr}_2\text{O}_3$  powders. *Phys. Rev.* **130**, 986–988 (1963).
27. Filoti, G., Kuncser, V., Rosenberg, M., Schinzer, C. & Kemmler-Sack, S. Variable rate spin freezing and long range antiferromagnetic order in  $\text{Bi}_2\text{FeRhO}_6$ . *J. Alloy. Comp.* **256**, 86–91 (1997).

**Acknowledgements** This work was supported by the EPSRC under EP/H000925/1. M.J.R. is a Royal Society Research Professor.

**Author Contributions** M.J.R. and J.B.C. developed the concept. P.M. carried out the materials synthesis, characterization and physical property measurements and analysis, H.N. performed the physical property measurements, M.J.P. and J.B.C. performed the structural analysis, J.A. analysed the magnetic and magnetoelectric data, P.B. built the magnetoelectric measurement equipment, P.S. performed and analysed the Mössbauer experiments. P.M. and M.J.R. wrote the first draft, all authors contributed to the development of the manuscript and to discussion as the project developed.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.J.R. (m.j.rosesinsky@liv.ac.uk) or J.B.C. (j.b.claridge@liv.ac.uk).

## METHODS

**Sample preparation.** Powder samples of  $(1-x)\text{BiTi}_{(1-y)/2}\text{Fe}_y\text{Mg}_{(1-y)/2}\text{O}_3-(x)\text{CaTiO}_3$ , in the compositional range  $x = 0.15, 0.60 \leq y \leq 0.90$ , were synthesized by a conventional solid-state reaction. The binary oxides  $\text{Bi}_2\text{O}_3$  (99.99% Alfa Aesar, pre-dried at 473 K),  $\text{CaCO}_3$  (99.997% Alfa Aesar, pre-dried at 473 K),  $\text{Fe}_2\text{O}_3$  (99.998% Alfa Aesar, pre-dried at 473 K),  $\text{TiO}_2$  (99.995% Alfa Aesar, pre-dried at 473 K) and  $\text{MgCO}_3 \cdot \text{Mg}(\text{OH})_2 \cdot x\text{H}_2\text{O}$  ( $x \approx 3$ , 99.995% Alfa Aesar, used as received) were weighed in stoichiometric amounts and ball milled in ethanol for 20 h. The mixtures obtained after evaporating ethanol were pelletized and calcined at 1,208 K for 12 h in a platinum-lined alumina crucible. These pellets were then re-ground thoroughly and re-pelletized, and subjected to a second calcination at 1,213 K for 12 h in platinum-lined alumina crucibles. The resulting powders were found to contain only the target phase with no minority phases visible by PXRD. Dense pellets (>95% of crystallographic density) suitable for property measurements were produced from these powders by the following protocol. First, 2 wt% polyvinyl butyral binder and 0.2 wt%  $\text{MnO}_2$  were added to the samples, and this mixture was ball-milled for 20 h. Second, the resultant mixture was pelletized (8 mm diameter) with a uniaxial press, followed by pressing at about  $2 \times 10^8$  Pa in a cold isostatic press. Third, these pellets were loaded into a platinum-lined alumina boat. Finally, a programmable tube furnace was used to heat the reaction under flowing oxygen to 943 K for 1 h, followed by 1,228 K for 3 h and 1,173 K for 12 h before cooling to room temperature at  $5 \text{ K min}^{-1}$ . The resultant pellets were found to contain no minority phases by PXRD. Their densities were measured using an Archimedes balance.

**Powder X-ray diffraction (PXRD).** All data were collected using a PANalytical X'Pert Pro diffractometer in Bragg-Brentano geometry with a monochromated  $\text{Co K}\alpha_1$  source (wavelength  $\lambda = 1.78896 \text{ \AA}$ ) and position-sensitive X'Celerator detector. Each sample was contained in a back-filled sample holder and rotated during the measurement. A programmable divergence slit was used to provide a constant illuminated area throughout the angular range. Data were collected in the angular range  $5^\circ \leq 2\theta \leq 130^\circ$  in steps of  $0.0167^\circ$ . Pawley refinements were carried out using the software package Topas Academic (version 5). For each PXRD pattern, background was modelled using a Chebyshev polynomial function with 12 refined parameters. Lattice parameters, a sample height correction, peak profile functions and model-independent peak intensities were refined. Peak profiles were modelled with a modified Thompson-Cox-Hastings pseudo-Voigt function. When fitting data to a single phase ( $R3c$  or  $P11a$  cells), a Stephens anisotropic strain broadening function was refined. In two-phase ( $R3c + Pna2_1$ ) refinements, this function was refined only for the rhombohedral ( $R3c$ ) phase.

**Electrical measurements.** For electric poling, gold was sputtered on both sides of thin disks (thickness of 130–160  $\mu\text{m}$  with tolerance of 10  $\mu\text{m}$ ). For  $P(E)$  and PUND measurements, silver conductive paint (RS Components) was applied on both sides of thin disks and cured at 393 K for 10 min. The edges were bevelled by approximately 0.2 mm to avoid electrical breakdown. The area of the electrode was measured under an optical microscope equipped with a camera and measurement software. The disk was loaded in a Radiant high-voltage test fixture. Silicone oil was used as a dielectric medium to avoid air breakdown.  $P(E)$  measurements were conducted using a Radiant ferroelectric tester system and an aixACCT piezoelectric evaluation system (aixPES). PUND measurements were carried out using the Radiant ferroelectric tester system with a square electric field pulse with a delay of 500 ms and pulse widths of 5 ms ( $x = 0.15, y = 0.60$ ) and 8 ms ( $x = 0.15, y = 0.80$ ). The remanent polarizations for positive ( $dP/2$ ) and negative ( $-dP/2$ ) applied electric fields are calculated as  $dP/2 = (P^* - P^+)/2$  and  $-dP/2 = (-P^* - (-P^+))/2$ , respectively, where  $P^*$  contains both remanent and non-remnant polarization, whereas  $P^+$  contains only the non-remnant polarization.  $P_r^*$  and  $P_r^+$  are equivalent polarizations of  $P^*$  and  $P^+$ , respectively, measured when the applied electric field is reduced to zero following the pulse.

**Leakage current measurements.** Leakage current was measured in an aixPES instrument using a triangular waveform in steps of 25 V and with a step duration of 2 s. A switching prepolarization pulse was applied before actual measurements.

**Resistivity measurements.** Resistivity was measured using the two-probe method in a Magnetic Property Measurement System (MPMS) XL-7 SQUID magnetometer (Quantum Design). The pellet was loaded into a modified dc-SQUID probe and connected to a Keithley 6430 sub-femtoamp remote sourcemeter.

**Impedance measurements.** Impedance and phase angles were measured using an Agilent LCR meter E4980 by applying an ac voltage of 0.5 V in the frequency range 20 Hz to 2 MHz.

**Magnetic measurements.** Magnetic measurements were carried out using MPMS XL-7 and MPMS3 systems (Quantum Design). For this, powder or pellet samples

were loaded into a polycarbonate capsule and fixed into a straight plastic drinking straw and then loaded into a dc-SQUID probe. The Néel temperature ( $T_N$ ) was determined from peak of  $dM_{\text{TRM}}/dT$ . The isothermal dc magnetization data were decomposed using the general function  $M(H) = \sum m_i(H)$ , where  $m_i$  are generic functions describing single magnetic components taking the form  $m_i(H) = a \tanh(\frac{H-b}{c}) + d$ . Here  $a$  represents the saturation magnetization,  $b$  the coercive field,  $c$  is a parameter that describes the squareness of the loop, and  $d$  is a linear term that includes paramagnetic, diamagnetic and antiferromagnetic contributions for the individual magnetic component. Above the magnetic ordering temperature of the perovskite phase, only one component was used to describe the isothermal magnetization assigned to a high Fe content impurity. Below the perovskite magnetic ordering temperature, two components were used.

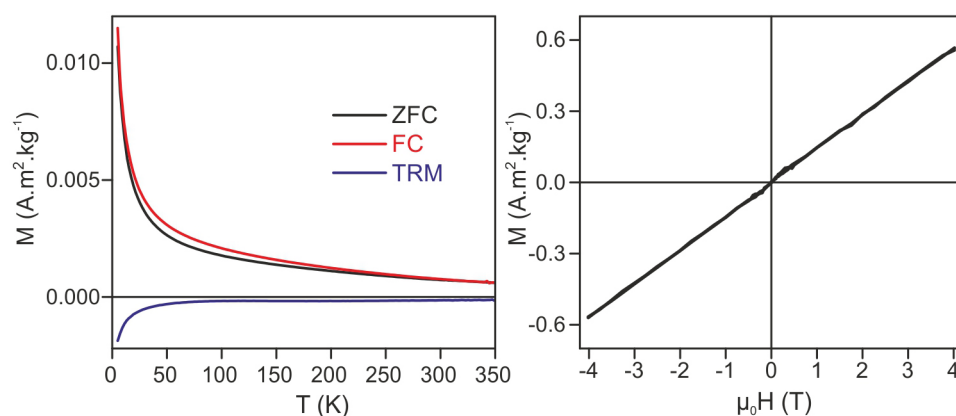
**Mössbauer spectroscopy.** Absorption mode Mössbauer spectroscopy measurements were performed at room temperature, using an electromagnetic Doppler drive system, a  $^{57}\text{Co}(\text{Rh})$   $\gamma$ -ray source with an actual activity of about 20 mCi and a Xe-gas Reuter Stokes proportional counter, and Canberra amplification, discrimination and scaling electronics. Samples were diluted with sucrose (icing sugar), for measurements at an approximate ratio of 0.2, to prevent excessive line-shape distortion and non-resonant absorption, owing to the high bismuth content of the samples. Custom modelling and nonlinear least-squares error minimization routines were used for the extraction of the spectroscopic parameters. Isomer shifts are reported with respect to the source.

**Magnetoelectric measurements.** Details of the magnetoelectric measurements set-up<sup>28</sup> and protocol<sup>28</sup> are described elsewhere. Note that the load resistance, mentioned and used in ref. 28 to gain a suitable voltage drop from the MPMS ac coil power supply, was omitted in our experiments; instead, the MPMS ac coil power supply was directly connected to the input of a Krohn-Hite 7600 M wide-band power amplifier. In this experiment, a sinusoidal electric field  $E = E_{\text{ac}} \cos(\omega t)$  (where  $\omega = 2\pi f$ , with  $f$  frequency, and  $E_{\text{ac}}$  is the electric field amplitude) is applied across the disk and the first harmonic of the complex ac magnetic moment,  $m(t) = (m' - im'') \cos(\omega t)$  is measured. The measurements were performed in the absence of any dc magnetic or electric fields. In this scenario, the real part of the electrically induced magnetic moment<sup>28</sup> is  $m' = \alpha E_{\text{ac}} \frac{V}{\mu_0}$ , where  $V$  is the sample volume. This moment involves only the linear magnetoelectric ( $\alpha$ ) effect; the higher-order effects are zero. The corresponding electrically induced volume ac magnetization is defined as  $M_{\text{ac}} = m'/V$ . To demonstrate the linear magnetoelectric effect on  $x = 0.15, y = 0.60$  and  $x = 0.15, y = 0.80$ , the electric field amplitude  $E_{\text{ac}}$  was varied and the induced moment was recorded. Linear magnetoelectric susceptibility ( $\alpha$ ) was calculated from a plot of volume ac magnetization amplitude

$M_{\text{ac}} (= m'/V)$  versus  $E_{\text{ac}}$ :  $\alpha = \mu_0 \frac{\Delta M_{\text{ac}}}{\Delta E_{\text{ac}}}$  (ref. 29). All measurements were performed at  $f = 1 \text{ Hz}$  with 20 blocks to average and 10 scans per measurement. The sensitivity of the experimental set-up used here is  $|m'| = V \times M_{\text{ac}} > 5 \times 10^{-12} \text{ A m}^{-2}$ . Prior to magnetoelectric measurements, disks were poled externally using the aixPES instrument at a field of  $100 \text{ kV cm}^{-1}$  for 15 min from 343 K to room temperature. Disks were then loaded into a modified dc-SQUID probe at 300 K and subjected to a magnetic field of 2 T for 30 min. After the removal of the electric and magnetic fields, electrodes were short circuited for 15 min before conducting magnetoelectric measurements at 300 K. For magnetoelectric measurements at 10 K and 150 K (for  $x = 0.15, y = 0.60$ ), the sample was cooled down to the measurement temperature in the presence of an electric field ( $3.5 \text{ kV cm}^{-1}$ ) and a magnetic field (2 T) and the protocol for 300 K measurement was followed. To determine the temperature dependence of  $\alpha$ , an electric field ( $3.5 \text{ kV cm}^{-1}$  for  $y = 0.60$  and  $2.7 \text{ kV cm}^{-1}$  for  $y = 0.80$ ) and a magnetic field of 2 T were applied at 300 K, followed by cooling to 10 K at a rate of  $1 \text{ K min}^{-1}$ ; the data were collected at 1 Hz. The temperature was stabilized for 5 min at each step before measurement. The room-temperature bulk dc resistivity of  $x = 0.15, y = 0.60$  is  $3.3 \times 10^{12} \Omega \text{ cm}$ , and that of  $x = 0.15, y = 0.80$  is  $2.1 \times 10^{12} \Omega \text{ cm}$ . The leakage currents observed for  $y = 0.60$  and  $y = 0.80$  are 0.35 nA (320 K) and 11.4 nA (360 K), respectively, at the maximum measurement fields. These values are too low to cause any artefacts in the magnetoelectric measurements. The upper limit of temperature in this measurement set-up is 360 K.

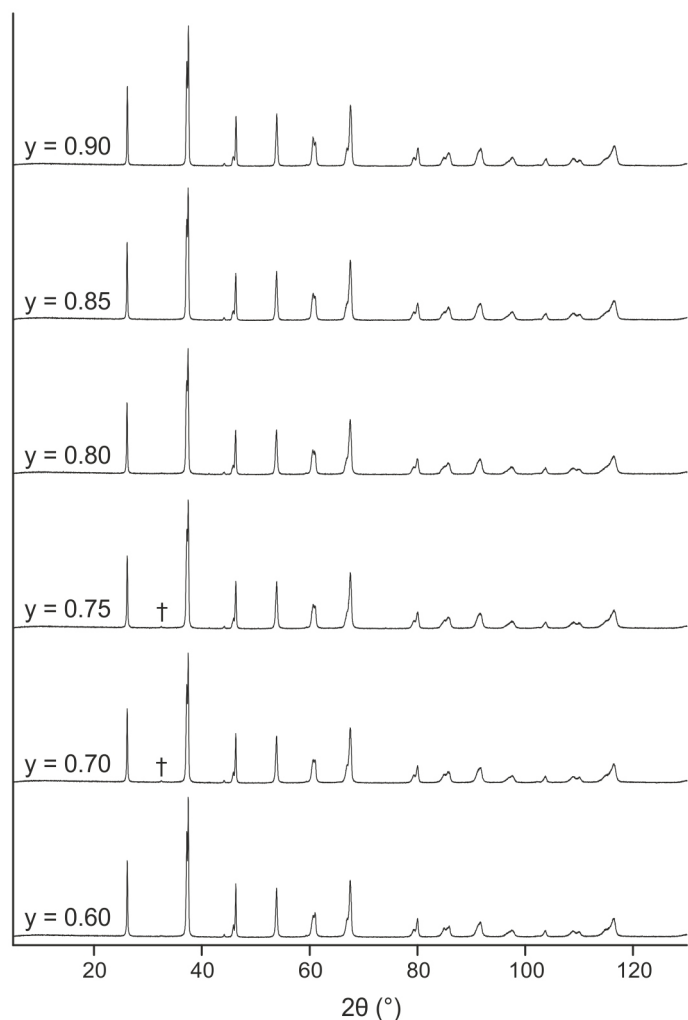
28. Borisov, P., Hochstrat, A., Shvartsman, V. V. & Kleemann, W. Superconducting quantum interference device setup for magnetoelectric measurements. *Rev. Sci. Instrum.* **78**, 106105 (2007).
29. Schmid, H. Some symmetry aspects of ferroics and single phase multiferroics. *J. Phys. Cond. Matter* **20**, 434201 (2008).



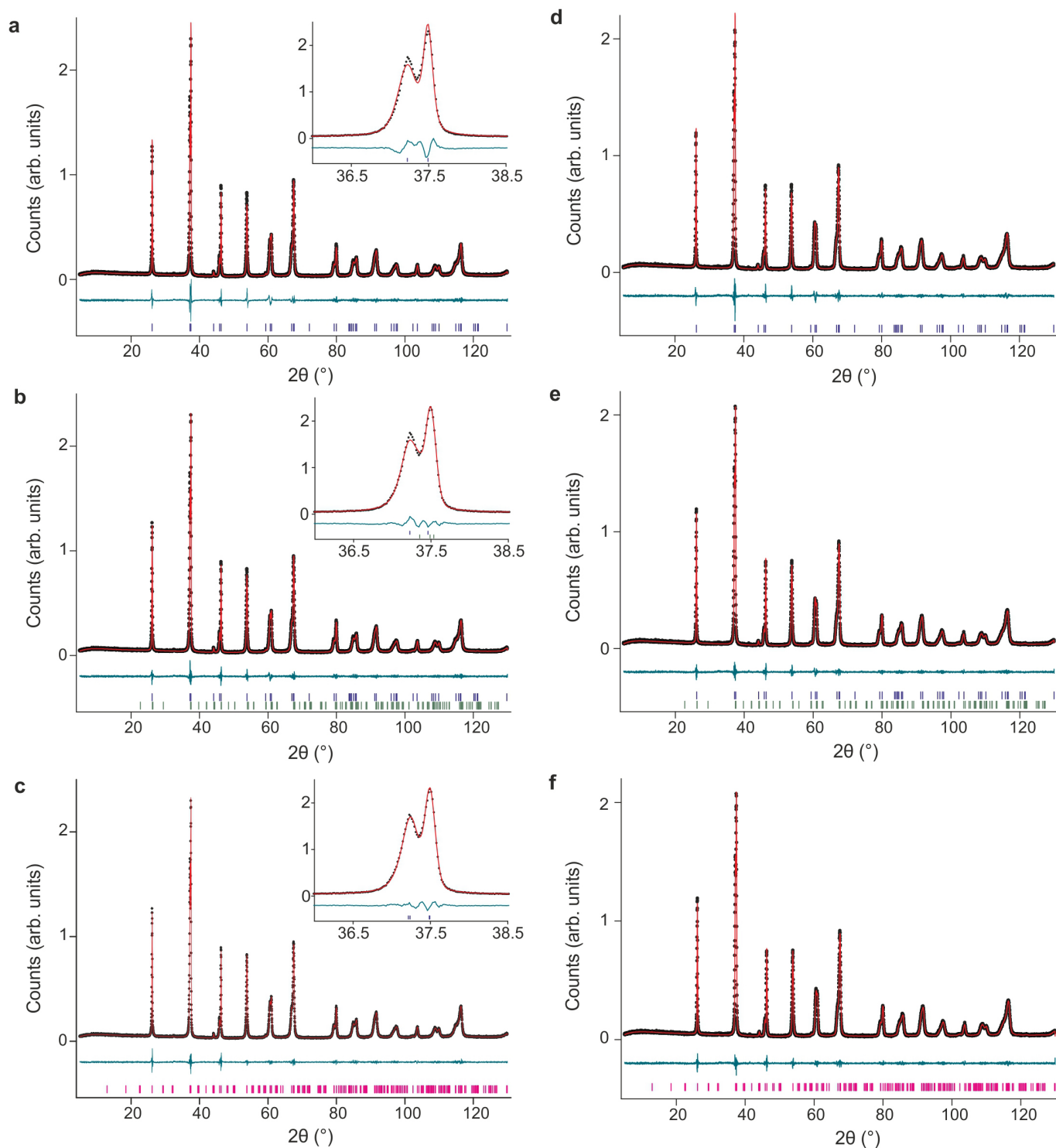


**Extended Data Figure 1 | Magnetic properties of composition  $x = 0.15$ ,  $y = 0.25$ .** Left, magnetization versus temperature, cooled in zero applied field (ZFC, black line), cooled in 1 mT applied field (FC, red line) and the thermal

remanent magnetization in zero applied field (TRM, blue line). Note negative TRM curve is due to a negative remanent magnetic field in the superconducting magnet. Right, magnetization versus magnetic field at 100 K.



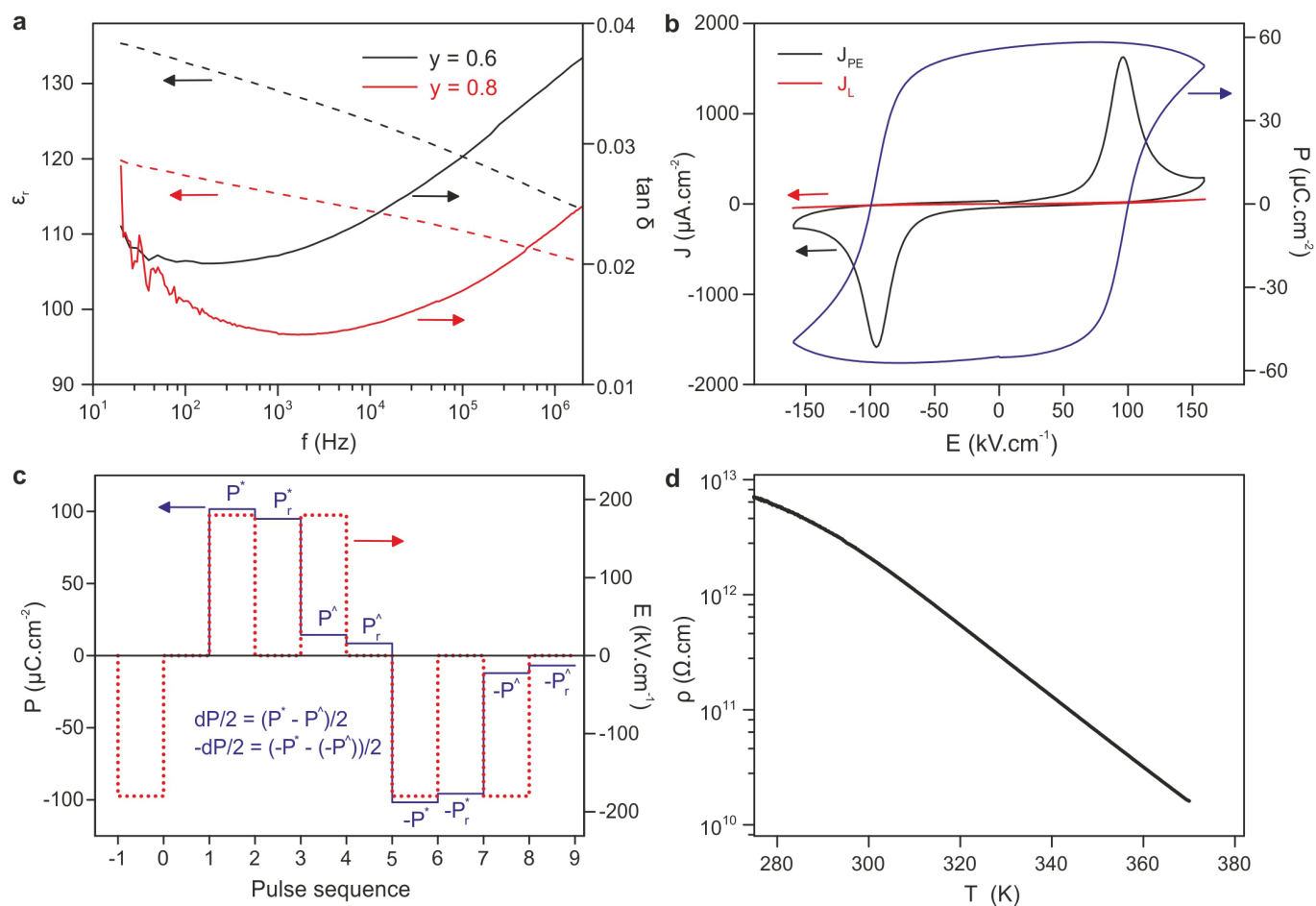
**Extended Data Figure 2 |** PXRD patterns obtained from six compositions of the series  $(1-x)\text{BiTi}_{(1-y)/2}\text{Fe}_y\text{Mg}_{(1-y)/2}\text{O}_3-(x)\text{CaTiO}_3$  where  $x = 0.15$ ,  $0.60 \leq y \leq 0.90$ . The weak reflection marked with the † symbol, which is visible in the  $y = 0.70$  and  $y = 0.75$  patterns, corresponds to the most intense reflection of sillenite ( $\text{Bi}_{25}\text{FeO}_{40}$ ). All other peaks are indexed to the target perovskite phase using rhombohedral, rhombohedral + orthorhombic, or monoclinic cells, as discussed in the text.



**Extended Data Figure 3 | Pawley fits to PXRD patterns collected from two compositions of the series  $(1-x)\text{BiTi}_{(1-y)/2}\text{Fe}_y\text{Mg}_{(1-y)/2}\text{O}_3-(x)\text{CaTiO}_3$ . a–f,  $x = 0.15$ ,  $y = 0.60$  (a–c) and  $x = 0.15$ ,  $y = 0.80$  (d–f) modelled as a single rhombohedral phase in space group  $R3c$  (a, d), as a combination of rhombohedral ( $R3c$ ) and orthorhombic ( $Pna2_1$ ) phases (b, e) and as a single**

**monoclinic phase in space group  $P11a$ , which is a subgroup of  $R3c$  and  $Pna2_1$  (c, f). Black circles,  $y_{\text{obs}}$ ; red line,  $y_{\text{calc}}$ ; teal line,  $(y_{\text{obs}} - y_{\text{calc}})$ ; blue markers,  $hkl$  ( $R3c$ ) reflections; green markers,  $hkl$  ( $Pna2_1$ ) reflections; magenta markers,  $hkl$  ( $P11a$ ) reflections. Insets are zooms of the main plots.**

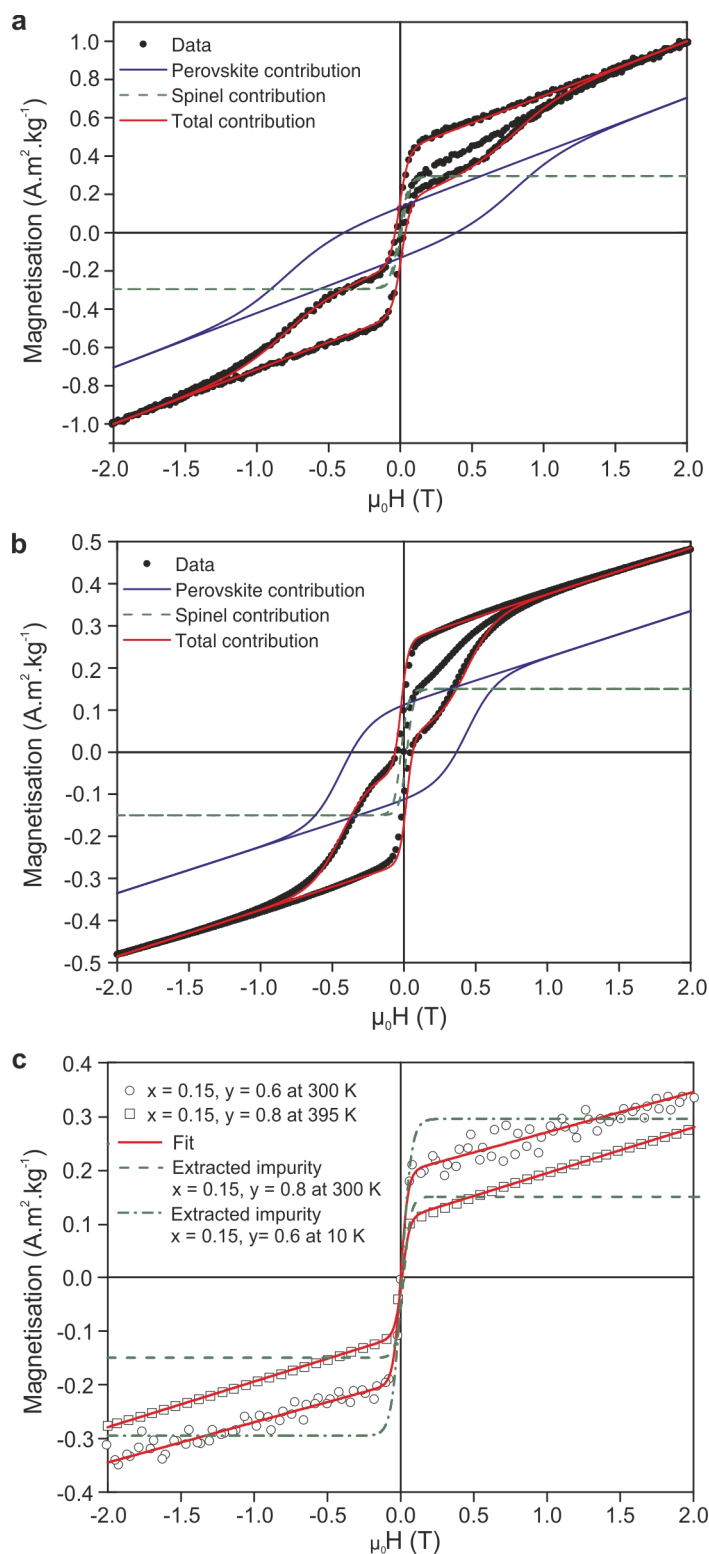




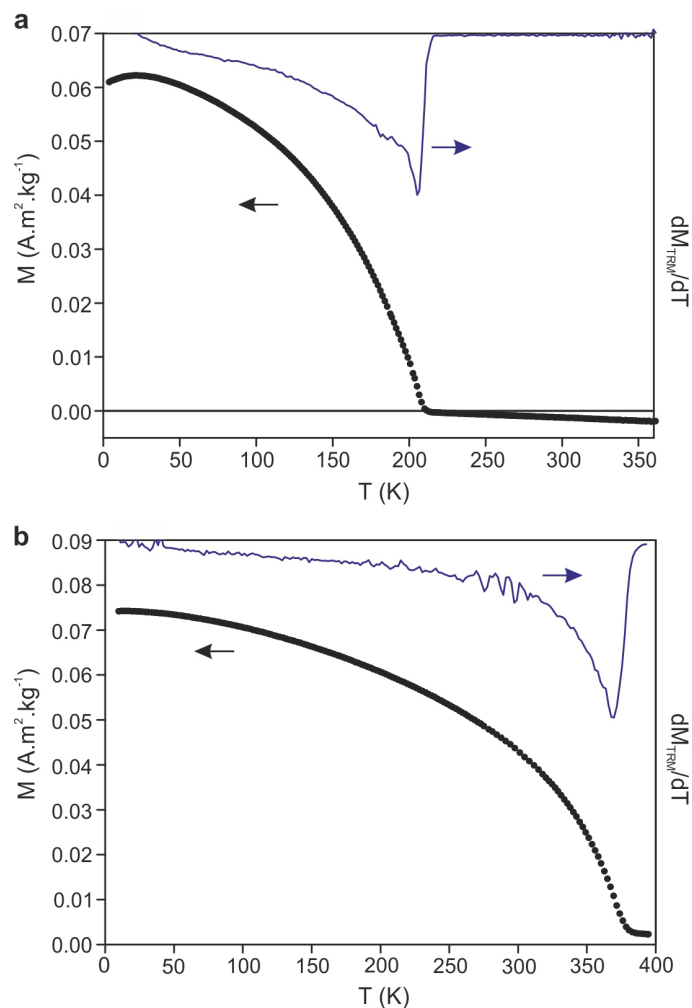
#### Extended Data Figure 4 | Dielectric, polarization and leakage

**characteristics.** **a**, Frequency dependence of dielectric permittivity (left axis, dashed line) and loss (right axis, solid line) at 300 K for  $x = 0.15$ ,  $y = 0.60$  (black) and  $x = 0.15$ ,  $y = 0.80$  (red). **b**, A typical  $P(E)$  loop (right axis, blue line) with the corresponding current density ( $J_{PE}$ ; left axis, black line) and the leakage current density ( $J_L$ ; left axis, red line) for  $x = 0.15$ ,  $y = 0.80$ . **c**, The

polarization (blue line, left axis) and electric field profile (red dotted line, right axis) from PUND measurement of  $x = 0.15$ ,  $y = 0.80$  (see Methods for details). **d**, Temperature dependence of dc resistivity of  $x = 0.15$ ,  $y = 0.80$ , showing highly insulating behaviour. In **a–c**, the arrows point to the relevant axis for each curve.



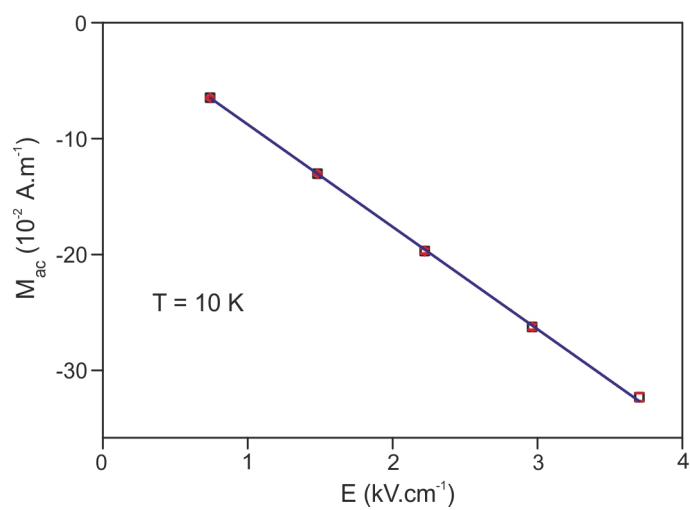
**Extended Data Figure 5 | Isothermal magnetization  $M(H)$ .** **a, b,**  $x = 0.15$ ,  $y = 0.60$  at  $T = 10 \text{ K} < T_N$  (**a**) and  $x = 0.15$ ,  $y = 0.80$  at  $T = 300 \text{ K} < T_N$  (**b**). The experimental data are represented as black filled circles. Red lines show the sum of the perovskite phase (blue line) and spinel impurity phase (green dashed line) contributions. **c,**  $x = 0.15$ ,  $y = 0.60$  at  $T = 300 \text{ K} > T_N$  and  $x = 0.15$ ,  $y = 0.80$  at  $T = 395 \text{ K} > T_N$ . The experimental data are represented as open circles ( $x = 0.15$ ,  $y = 0.60$ ) or squares ( $x = 0.15$ ,  $y = 0.80$ ); green dash-dotted and dashed lines show extracted spinel impurity contributions for  $x = 0.15$ ,  $y = 0.60$  and  $x = 0.15$ ,  $y = 0.80$ , respectively; red lines show fits to the data.



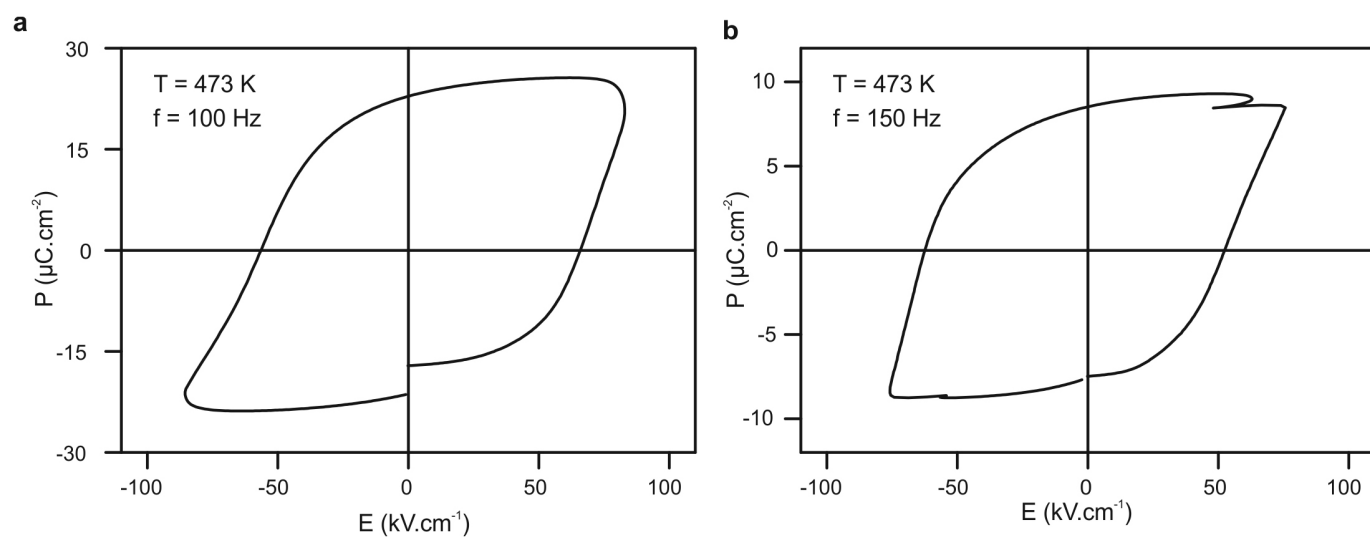
**Extended Data Figure 6 | Thermal remanent magnetization data.**

**a, b,** Thermal remanent magnetization (TRM; left axis, black circles) and derivative of TRM with respect to temperature ( $dM_{TRM}/dT$ ; right axis, blue lines) for  $x = 0.15, y = 0.60$  (a) and  $x = 0.15, y = 0.80$  (b). Arrows indicate the axis that each dataset corresponds to.





**Extended Data Figure 7 | Linear magnetoelectric effect for  $x = 0.15$ ,  $y = 0.60$  at 10 K.** Red squares are mean values, error bars in red are standard errors from 10 repeated measurements. The blue line is a linear fit to the data.



**Extended Data Figure 8** |  $P(E)$  measurements above room temperature. **a, b**, Measurements for  $x = 0.15$ ,  $y = 0.60$  at frequency  $f = 100\text{ Hz}$  (**a**) and  $x = 0.15$ ,  $y = 0.80$  at frequency  $f = 150\text{ Hz}$  (**b**) at  $473\text{ K}$ .

Extended Data Table 1 | Refined lattice parameters and agreement factors from Pawley fits to PXRD data

Composition	Refined Lattice Parameters (space group <i>P11a</i> )					Weighted Profile R-factor ( $R_{wp}$ )			Goodness of fit ( $\chi^2$ )		
	<i>a</i> (Å)	<i>b</i> (Å)	<i>c</i> (Å)	$\gamma$ (°)	Volume (Å <sup>3</sup> )	<i>R3c</i>	<i>R3c</i> + <i>Pna2</i> <sub>1</sub>	<i>P11a</i>	<i>R3c</i>	<i>R3c</i> + <i>Pna2</i> <sub>1</sub>	<i>P11a</i>
<b>x = 0.15, y = 0.60</b>	5.6037(3)	7.9047(6)	5.5666(1)	89.433(7)	246.56(2)	7.421	6.373	6.169	2.104	1.583	1.499
<b>x = 0.15, y = 0.80</b>	5.6019(7)	7.903(1)	5.5641(1)	89.40(1)	246.32(5)	6.883	6.136	5.969	1.821	1.477	1.392

Refined lattice parameters (*a*, *b*, *c*,  $\gamma$ ) and the corresponding unit cell volumes, obtained by fitting to a  $\sqrt{2}a_p \times 2a_p \times \sqrt{2}a_p$  unit cell in space group *P11a*, and agreement factors ( $R_{wp}$  and  $\chi^2$ ) from Pawley fits to PXRD data, fitted in three different candidate space groups, for compositions *x* = 0.15, *y* = 0.60 and *x* = 0.15, *y* = 0.80.



**Extended Data Table 2 | Spectroscopic parameters from Mössbauer data fitting of  $x = 0.15$ ,  $y = 0.80$  at 300 K**

Component	Area (%)	$\delta$ (mm.s <sup>-1</sup> )	Q (mm.s <sup>-1</sup> )	$B_{\text{hf}}$ (T)
1	30.8(1)	0.29(5)	0.033(7)	46(1)
2	67.8(1)	0.31(5)	-0.003(2)	23(4)
3	0.6(1)	0.30(5)	-0.063(4)	33.5(5)
4	0.7(1)	0.30(5)	-0.32(1)	41.7(5)

The area, isomer shift ( $\delta$ ), electric quadrupole moment ( $Q$ ) and hyperfine field ( $B_{\text{hf}}$ ) for different components, extracted from a multicomponent fit, with standard errors in parentheses.

# The contribution of outdoor air pollution sources to premature mortality on a global scale

J. Lelieveld<sup>1,2</sup>, J. S. Evans<sup>3,4</sup>, M. Fnais<sup>5</sup>, D. Giannadaki<sup>2</sup> & A. Pozzer<sup>1</sup>

Assessment of the global burden of disease is based on epidemiological cohort studies that connect premature mortality to a wide range of causes<sup>1–5</sup>, including the long-term health impacts of ozone and fine particulate matter with a diameter smaller than 2.5 micrometres (PM<sub>2.5</sub>)<sup>3–9</sup>. It has proved difficult to quantify premature mortality related to air pollution, notably in regions where air quality is not monitored, and also because the toxicity of particles from various sources may vary<sup>10</sup>. Here we use a global atmospheric chemistry model to investigate the link between premature mortality and seven emission source categories in urban and rural environments. In accord with the global burden of disease for 2010 (ref. 5), we calculate that outdoor air pollution, mostly by PM<sub>2.5</sub>, leads to 3.3 (95 per cent confidence interval 1.61–4.81) million premature deaths per year worldwide, predominantly in Asia. We primarily assume that all particles are equally toxic<sup>5</sup>, but also include a sensitivity study that accounts for differential toxicity. We find that emissions from residential energy use such as heating and cooking, prevalent in India and China, have the largest impact on premature mortality globally, being even more dominant if carbonaceous particles are assumed to be most toxic. Whereas in much of the USA and in a few other countries emissions from traffic and power generation are important, in eastern USA, Europe, Russia and East Asia agricultural emissions make the largest relative contribution to PM<sub>2.5</sub>, with the estimate of overall health impact depending on assumptions regarding particle toxicity. Model projections based on a business-as-usual emission scenario indicate that the contribution of outdoor air pollution to premature mortality could double by 2050.

Air pollution is associated with many health impacts, including chronic obstructive pulmonary disease (COPD) linked to enhanced ozone (O<sub>3</sub>), and acute lower respiratory illness (ALRI), cerebrovascular disease (CEV), ischaemic heart disease (IHD), COPD and lung cancer (LC) linked to PM<sub>2.5</sub> (ref. 8). Many previous studies have been based on air quality measurements, largely focusing on urban pollution<sup>3,4,11–14</sup>. Atmospheric chemistry and transport models have been used to account for other environments, including those for which no measurement data are available<sup>15–22</sup>.

Recently, enhanced resolution regional and global models and satellite data have been applied to improve estimates of PM<sub>2.5</sub> and O<sub>3</sub> concentrations and their impact on air quality<sup>19–24</sup>. Here we present results obtained with an atmospheric chemistry–general circulation model, applied at high resolution to compute global air quality changes, combined with population data, country-level health statistics and pollution exposure response functions (Methods). Our calculations of air pollution related mortality are based on the method of the global burden of disease (GBD) for 2010 (ref. 5), applying improved exposure response functions that more realistically account for health effects at very high PM<sub>2.5</sub> concentrations compared to former assessments<sup>8</sup>. This is particularly relevant for some parts of the world where air pollution has increased nearly unabated and for future scenarios that project the continued growth of emissions. Following the

GBD<sup>5</sup> we also include desert dust (which is largely natural) with PM<sub>2.5</sub>; hence strictly speaking we assess the effects of atmospheric composition.

The air quality guidelines of the World Health Organization (WHO) and national regulatory policies are based on exposure response functions that rely on PM<sub>2.5</sub> mass concentrations, implicitly treating all fine particles as equally toxic without regard to their source and chemical composition. However, expert elicitation suggests that carbonaceous particles are more toxic than crustal material, nitrates and sulfates<sup>10</sup>. A recent study<sup>25</sup> finds that PM<sub>2.5</sub> from coal combustion leads to increased mortality risk from cardiovascular disease and LC, but that the evidence is much weaker for other sources, whereas estimates using non-specific PM<sub>2.5</sub> mass alone may underestimate the total effect of PM<sub>2.5</sub> on mortality. Further, this study did not find support for mortality from biomass combustion and soil dust particles<sup>25</sup>. However, this and a subsequent report by the Health Effects Institute in the USA also note that there were only a limited number of cities in these investigations where these sources and components were likely to be measured consistently<sup>26,27</sup>. While the evidence for differential toxicity is far from conclusive, we conducted a secondary analysis assuming that carbonaceous PM<sub>2.5</sub> is five times more toxic than inorganic particles, though maintaining the same overall health impact of PM<sub>2.5</sub>.

We have calculated premature mortality linked to CEV, COPD, IHD and LC for adults  $\geq 30$  years old, and ALRI for infants  $< 5$  years old (Table 1 and Extended Data Tables 1 and 2). Our estimate of the global PM<sub>2.5</sub> related mortality in 2010 is 3.15 million people with a 95% confidence interval (CI95) of 1.52–4.60 million. The main causes are CEV (1.31 million) and IHD (1.08 million), and secondary causes are COPD (374 thousand), ALRI (230 thousand) and LC (161 thousand). Our global estimate of O<sub>3</sub> related mortality by COPD is 142 (CI95: 90–208) thousand. Our total estimate of 3.30 (CI95: 1.61–4.81) million people in 2010 agrees closely with the GBD<sup>5</sup>. This is in addition to the estimated 3.54 million deaths per year caused by indoor air pollution due to use of solid fuels for cooking and heating<sup>6</sup>. Figure 1 shows the geographic distribution and demonstrates the locations of hotspots in China, India and many of the large urban centres.

Considering the global population of 6.8 billion in 2010, it follows that the mean per capita mortality attributable to air pollution is about 5 per 10,000 person-years. Of these 5 persons per 10,000 worldwide, about 2 die by CEV, 1.6 by IHD, 0.8 by COPD, 0.35 by ALRI and 0.25 by LC. The highest per capita mortality is found in the Western Pacific region, followed by the Eastern Mediterranean and Southeast Asia. The combination of high per capita mortality with high population density explains the (by far) highest number of deaths in the Western Pacific, China being the main contributor (1.36 million per year). Note that the mortality attributable to air pollution in China is approximately an order of magnitude higher than that attributable to Chinese road transport injuries and HIV/AIDS, and ranks among the top causes of death<sup>28</sup>. Southeast Asia has the second highest premature mortality, where India is the main contributor (0.65 million per year). The global

<sup>1</sup>Max Planck Institute for Chemistry, Atmospheric Chemistry Department, 55128 Mainz, Germany. <sup>2</sup>The Cyprus Institute, Energy, Environment and Water Research Center, 1645 Nicosia, Cyprus. <sup>3</sup>Harvard School of Public Health, Boston, Massachusetts 02215, USA. <sup>4</sup>Cyprus International Institute for Environment and Public Health, Cyprus University of Technology, 3041 Limassol, Cyprus. <sup>5</sup>King Saud University, College of Science, Riyadh 11451, Saudi Arabia.

**Table 1 | Premature mortality related to PM<sub>2.5</sub> and O<sub>3</sub> for the population <5 and ≥30 years old**

WHO region	Year	Population (×10 <sup>6</sup> )	Mortality attributable to air pollution (deaths × 10 <sup>3</sup> )						Total
			PM <sub>2.5</sub>			O <sub>3</sub>			
			ALRI < 5 yr	IHD ≥ 30 yr	CEV ≥ 30 yr	COPD ≥ 30 yr	LC ≥ 30 yr	COPD ≥ 30 yr	
Africa	2010	809	90	55	77	11	2	2	237
	2050	1,807	158	185	262	38	5	12	660
Americas	2010	930	0	44	8	4	7	5	68
	2050	1,191	0	75	15	7	11	11	119
Eastern Mediterranean	2010	602	56	115	86	12	5	12	286
	2050	1,021	66	321	246	37	13	40	723
Europe	2010	867	1	239	95	13	27	6	381
	2050	886	1	307	156	18	37	11	530
Southeast Asia	2010	1,762	64	327	250	124	15	82	862
	2050	2,332	104	865	807	419	48	227	2,470
Western Pacific	2010	1,812	19	299	794	209	107	35	1,463
	2050	1,861	16	413	1,120	309	155	57	2,070
World	2010	6,783	230	1,079	1,311	374	161	142	3,297
	2050	9,098	346	2,166	2,604	828	270	358	6,572

Regions are defined by the World Health Organization, see Extended Data Table 1. Results for 2050 are based on a business-as-usual scenario.

mortality linked to air pollution is strongly influenced by these high numbers in Asia.

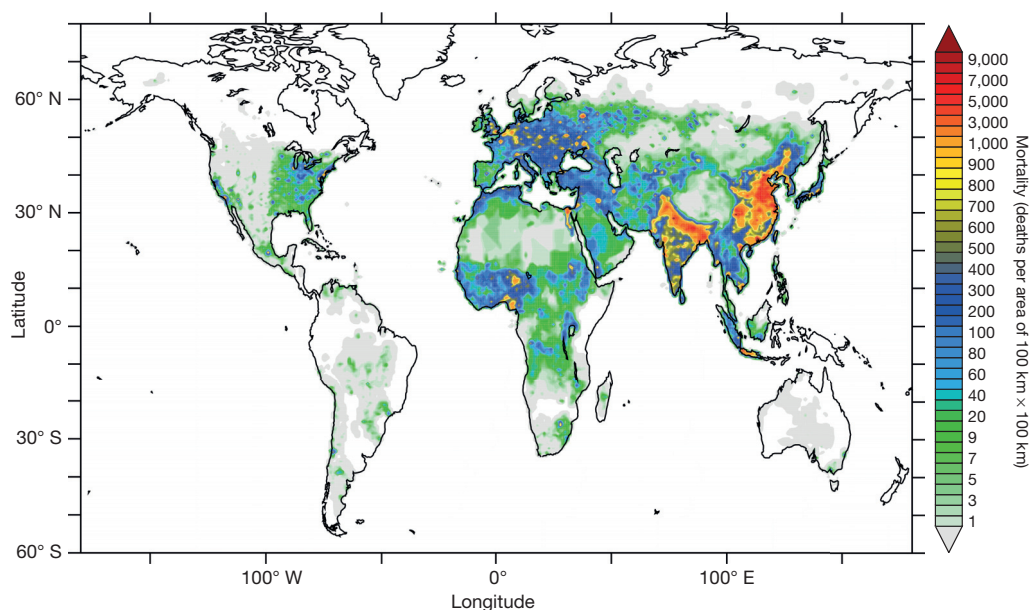
We determined the impacts of seven source categories by subtracting them one by one from the emissions in our model. These sensitivity calculations show the efficacy of individually controlling these sources. The 15 countries with highest premature mortality attributable to air pollution in 2010 are listed in Table 2 along with the contribution of each source category. Residential and commercial energy use (RCO) is the largest source category worldwide, contributing nearly one-third, and almost a factor of 2 more under the alternative assumption of differential toxicity. Note that this only refers to mortality by outdoor exposure to this source. Our estimate of 1.0 million deaths per year by RCO is in addition to the 3.54 million deaths per year due to indoor air pollution from essentially the same source<sup>5</sup>.

The next largest anthropogenic source category is agriculture (AGR), contributing one-fifth; however, this reduces significantly under the assumption of differential particle toxicity. The successive principal anthropogenic categories are power generation (PG), industry (IND), biomass burning (BB) and land traffic (TRA), and taken together they cause nearly one-third of all air pollution mortality. If carbonaceous particles are five times more toxic than sulfates and nitrates, these sources together account for one-quarter of the mortal-

ity. Natural sources make up for the remaining one-sixth of the total. However, if crustal material is five times less toxic than carbonaceous PM<sub>2.5</sub> this reduces considerably. The most important source category in each region in 2010 is shown in Fig. 2.

RCO is foremost in the populous parts of Asia. It refers to small combustion sources, especially biofuel use (for heating and cooking), and also waste disposal and diesel generators. In China it contributes about 32%, in India, Bangladesh, Indonesia and Vietnam 50–60%, while in Nepal it is highest with nearly 70% (Extended Data Table 3). In western countries it is typically 5–10%, although in France and Poland it contributes about 15%. The contribution of this pollution source to mortality is sensitive to toxicity assumptions and large uncertainty related to IHD. Because of the comparatively large fraction of carbonaceous PM<sub>2.5</sub>, under our alternative calculations where these aerosols are five times more toxic, RCO increases from 31% to 59% of global air pollution mortality. If, on the other hand, we assume that RCO does not contribute to IHD mortality, this fraction decreases from 31% to 26% (Methods).

Agriculture (AGR) has a remarkably large impact on PM<sub>2.5</sub>, and is the leading source category in Europe, Russia, Turkey, Korea, Japan and the Eastern USA (Fig. 2). In many European countries, its contribution is 40% or higher. Agricultural releases of ammonia (NH<sub>3</sub>) from



**Figure 1 | Mortality linked to outdoor air pollution in 2010.** Units of mortality, deaths per area of 100 km × 100 km (colour coded). In the white areas, annual mean PM<sub>2.5</sub> and O<sub>3</sub> are below the concentration–response thresholds where no excess mortality is expected.



**Table 2 | Top 15 ranked countries of premature mortality linked to outdoor air pollution in 2010**

Country	Deaths ( $\times 10^3$ )	Residential energy	Agriculture	Natural	Power generation	Industry	Biomass burning	Land traffic
China	1,357	<b>32 (76)</b>	29 (7)	9 (3)	18 (7)	8 (3)	1 (2)	3 (2)
India	645	<b>50 (77)</b>	6 (1)	11 (1)	14 (5)	7 (3)	7 (9)	5 (4)
Pakistan	111	<b>31 (67)</b>	2 (1)	<b>57 (23)</b>	2 (1)	2 (2)	2 (3)	3 (3)
Bangladesh	92	<b>55 (78)</b>	10 (2)	0 (0)	15 (6)	7 (2)	7 (8)	6 (4)
Nigeria	89	14 (31)	1 (0)	<b>77 (52)</b>	0 (0)	0 (0)	8 (16)	0 (0)
Russia	67	7 (18)	<b>43 (26)</b>	1 (0)	22 (17)	8 (5)	8 (21)	11 (13)
USA	55	6 (12)	29 (17)	2 (2)	<b>31 (19)</b>	6 (5)	5 (9)	21 (36)
Indonesia	52	<b>60 (64)</b>	2 (0)	0 (0)	5 (3)	4 (2)	27 (29)	2 (2)
Ukraine	51	6 (13)	<b>52 (32)</b>	0 (0)	18 (17)	9 (7)	5 (18)	10 (13)
Vietnam	44	<b>51 (74)</b>	12 (2)	0 (0)	13 (4)	8 (3)	12 (14)	4 (3)
Egypt	35	1 (2)	3 (3)	<b>92 (88)</b>	2 (2)	1 (1)	0 (1)	1 (3)
Germany	34	8 (17)	<b>45 (26)</b>	0 (0)	13 (10)	13 (8)	1 (3)	20 (36)
Turkey	32	9 (20)	<b>29 (19)</b>	15 (6)	19 (14)	11 (8)	6 (19)	11 (14)
Iran	26	1 (3)	6 (6)	<b>81 (75)</b>	4 (4)	3 (3)	1 (2)	4 (7)
Japan	25	12 (29)	<b>38 (22)</b>	0 (0)	17 (15)	18 (14)	5 (8)	10 (12)
World	3,297	<b>31 (59)</b>	20 (7)	18 (11)	14 (7)	7 (3)	5 (8)	5 (5)

Columns 3–9 show contributions (%) of the seven main source categories, the leading one in bold. For details and additional countries, see Extended Data Table 3. In parentheses are shown sensitivity calculations with carbonaceous particles having a five times larger impact than inorganic aerosol compounds.

fertilizer use and domesticated animals affect air quality through several multiphase chemical pathways, forming ammonium sulphate and nitrate. Since  $\text{NH}_3$  abundance is often limiting in  $\text{PM}_{2.5}$  formation, reduction of its emissions can make an important contribution to air quality control<sup>29</sup>. As agricultural emissions mostly form inorganic  $\text{PM}_{2.5}$ , the impact on mortality diminishes under the assumption that carbonaceous  $\text{PM}_{2.5}$  is five times more toxic.

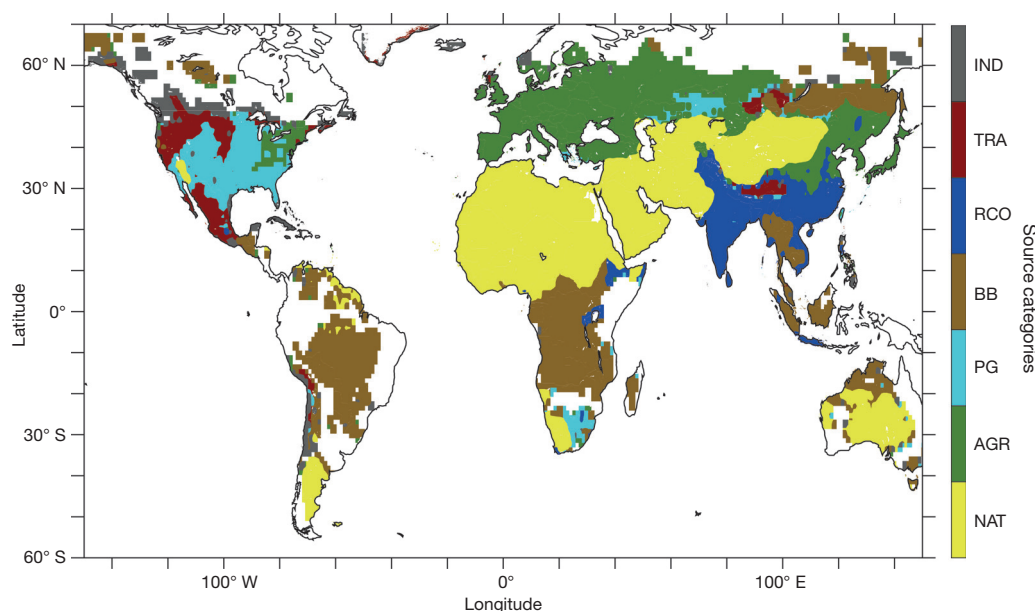
Natural sources (NAT) contribute strongly to mortality, being dominant in northern Africa and the Middle East, and also a leading category in Central Asia (Table 2 and Fig. 2). Although we categorize airborne desert dust as natural, a fraction is anthropogenic due to the role of humans in desertification and agricultural practices<sup>30</sup>. The chronic health and mortality impacts associated with exposure to dust are more uncertain than those due to typical air pollution in industrialized countries where most of the epidemiological cohort studies have been carried out. If all fine particles are equally toxic, then natural sources are responsible for about one-sixth of air pollution mortality. If fine carbonaceous particles are five times more toxic than crustal material, then natural sources account for only about one-tenth of air pollution induced mortality.

Power generation (PG) by fossil fuel fired power plants is the third largest anthropogenic source category, being an important source of

$\text{SO}_2$  and  $\text{NO}_x$ , which are converted to sulfate and nitrate in the atmosphere. It accounts for about one-seventh of population exposure to  $\text{PM}_{2.5}$  and  $\text{O}_3$ . Power plant emissions are quite important in the USA (>30%) and in Russia, Korea and Turkey (roughly 20%). Emissions from power generation also have particularly large impacts on fine particle concentrations in the Middle East, but frequently these go unnoticed as they are masked by desert dust. The role of this source is sensitive to the assumed  $\text{PM}_{2.5}$  toxicity, reducing by a factor of 2 if sulfate and nitrate are five times less toxic than carbonaceous  $\text{PM}_{2.5}$ .

Industry (IND) is among the smaller source categories, with a global fraction of about 7% (Table 2); nevertheless, it contributes about twice this percentage in most of the western world. It includes iron and steel, chemical, pulp and paper, food, solvent and other manufacturing sectors, oil refineries and fuel production. This source of air pollution is generally significant in industrialized countries and emerging economies, but rarely the leading cause of premature mortality. Under the differential toxicity assumption, its contribution to mortality would reduce by more than a factor of 2.

Our calculations suggest that land traffic (TRA) emissions are responsible for about one-fifth of mortality by ambient  $\text{PM}_{2.5}$  and  $\text{O}_3$  in Germany, the UK and the USA, while globally they account



**Figure 2 | Source categories responsible for the largest impact on mortality linked to outdoor air pollution in 2010.** Source categories (colour coded): IND, industry; TRA, land traffic; RCO, residential and commercial energy use (for example, heating, cooking); BB, biomass burning; PG, power generation; AGR, agriculture; and NAT, natural. In the white areas, annual mean  $\text{PM}_{2.5}$  is below the concentration–response threshold.

for about 5%. Because emissions of  $\text{NO}_x$  are the dominant source of traffic-related  $\text{PM}_{2.5}$  in the form of nitrate, together with carbonaceous  $\text{PM}_{2.5}$ , the results from our alternative calculations—assuming carbonaceous particles are five times more toxic than nitrates and other inorganics—also indicate a 5% contribution, globally. Note that this contribution is likely to be a lower limit as traffic also emits other pollutants that are not included or influential on  $\text{PM}_{2.5}$  (ref. 31) (Methods).

Biomass burning (BB) is also a relatively small source category with a global contribution of about 5%. Nevertheless, its areal range is large, for example in South America and Africa. It is the main source of air pollution in large parts of Canada, Siberia, Africa, South America and Australia. Because in many parts of these countries annual mean  $\text{PM}_{2.5}$  is below the concentration–response threshold (Methods), these areas are shown white in Fig. 2. Biomass burning is also widespread in southeastern Asia, although in populous parts of Vietnam and Indonesia (for example, Java) residential energy use is larger and therefore the leading category (Table 2).

In the Southern Hemisphere biomass burning is generally the leading contributor to  $\text{PM}_{2.5}$ , with some exceptions. In Brazil it contributes about 70%, and in many African countries its impact can also be high, up to >90% in Angola. Note that the health impacts of  $\text{PM}_{2.5}$  from biomass burning are quite uncertain, especially the attribution of IHD related mortality, due to a dearth of epidemiological cohort studies in regions where this pollution source predominates (Methods). Our calculations suggest that it is responsible for between 5% (equal toxicity) and 8% (differential toxicity) of air pollution induced mortality.

To understand how the premature mortality attributable to air pollution may develop in the coming decades, we applied a business-as-usual (BaU) emission scenario for the years 2025 and 2050, assuming that only currently agreed legislation is implemented that will affect future emissions<sup>32</sup>. Thus air quality and emission standards are fixed. Results for 2050 are presented here, and for 2025 in Extended Data Fig. 2 and Extended Data Tables 4, 5. Under the BaU scenario, moderate though significant increases of premature mortality will occur in Europe and the Americas, to a large degree in urban areas. Large increases are projected in Southeast Asia and the Western Pacific, leading to a global growth of premature mortality to 6.6 (CI95: 3.4–9.3) million (+100%) in 2050 (Table 1). This compares to a negligible population increase of infants (<5 years old), and a substantial increase (+68%) among people  $\geq 30$  years old in 2050 (implying an ageing population). Globally, the per capita mortality is projected to increase from 5 per 10,000 person-year in 2010 to about 7 per 10,000 person-year in 2050. The mortality attributable to air pollution will continue to be dominated by Asia with an unchanged fraction of about 75%.

The urban population is expected to grow relatively rapidly from 3.6 billion in 2010 to 5.2 billion in 2050, and combined with increasing air pollution concentrations the health impacts will escalate. Our estimate of urban premature mortality by outdoor air pollution in 2010 is 2.0 million, increasing to 4.3 million in 2050, representing 60% of the global total in 2010 and 65% in 2050. Urban population growth is responsible for part of this change, but the levels of air pollution in urban areas are also projected to grow rapidly. This is evident from our finding that the per capita mortality attributable to air pollution in 2010 is about 50% higher in urban than in rural environments. Under the BaU scenario this difference is expected to increase to nearly 90% in 2050.

Recently, much emphasis has been placed on rapidly emerging megacities (Methods). We calculate that 17 megacities and conurbations in Asia rank among the top 30 in terms of premature mortality worldwide, the leading one being the Pearl River Delta. When viewed instead from the perspective of individual risk, Tianjin and Beijing rank highest (Extended Data Table 6). While the per capita mortality attributable to air pollution is already extraordinary in Chinese megacities, according to the BaU scenario it will become even higher in Chinese and also Indian megacities by 2050. The combined premature

mortality in the 30 largest conurbations accounts for about 7% of the worldwide burden of air pollution, indicating the relevance of all urban areas.

Our results suggest that if the projected increase in mortality attributable to air pollution is to be avoided, intensive air quality control measures will be needed, particularly in South and East Asia. The poorly characterized uncertainty about the relative toxicity of various classes of particles such as sulfates, nitrates, organics, crustal materials, black carbon, and especially smoke from biomass combustion, limits unambiguous attribution of sources. Nevertheless, our study suggests that emissions from residential energy use should be considered in air pollution control strategies and, if all fine particles are equally toxic, the reduction of agricultural emissions would improve air quality. An improvement in the efficacy of air pollution controls requires a better understanding of the relative toxicity of particles from various emissions sources.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 10 May 2014; accepted 27 July 2015.

- Murray, C. & Lopez, A. D. *The Global Burden of Disease: A Comprehensive Assessment of Mortality and Disability from Diseases, Injuries, and Risk Factors in 1990 and Projected in 2020* (Harvard Univ. Press, 1996).
- Ezzati, M. *et al.* Selected major risk factors and global and regional burden of disease. *Lancet* **360**, 1347–1360 (2002).
- Ostro, B. *Outdoor Air Pollution: Assessing the Environmental Burden of Disease at National and Local Levels* (World Health Organization Environmental Burden of Disease Series No. 5, WHO, Geneva, 2004).
- Cohen, A. J. *et al.* The global burden of disease due to outdoor air pollution. *J. Toxicol. Environ. Health A* **68**, 1301–1307 (2005).
- Lim, S. S. *et al.* A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet* **380**, 2224–2260 (2012); correction **381**, 628 (2013).
- Pope, C. A. III & Dockery, D. W. Health effects of fine particulate air pollution: lines that connect. *J. Air Waste Manag. Assoc.* **56**, 709–742 (2006).
- Beelen, R. *et al.* Effects of long-term exposure to air pollution on natural-cause mortality: an analysis of 22 European cohorts within the multicentre ESCAPE project. *Lancet* **383**, 785–795 (2014).
- Burnett, R. T. *et al.* An integrated risk function for estimating the Global Burden of Disease attributable to ambient fine particulate matter exposure. *Environ. Health Perspect.* **122**, 397–403 (2014).
- Jerrett, M. *et al.* Long-term ozone exposure and mortality. *N. Engl. J. Med.* **360**, 1085–1095 (2009).
- Tuomisto, J. T., Wilson, A., Evans, J. S. & Tainio, M. Uncertainty in mortality response to airborne fine particulate matter: combining European air pollution experts. *Reliab. Eng. Syst. Saf.* **93**, 732–744 (2008).
- Pope, C. A. III *et al.* Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution. *J. Am. Med. Assoc.* **287**, 1132–1141 (2002).
- Prüss-Ustün, A., Bonjour, S. & Corvalan, C. The impact of the environment on health by country: a meta-synthesis. *Environ. Health* **7**, <http://dx.doi.org/10.1186/1476-069X-7-7> (2008).
- Russell, A. G. & Brunekreef, B. A focus on particulate matter and health. *Environ. Sci. Technol.* **43**, 4620–4625 (2009).
- Gurjar, B. R. *et al.* Human health risks in megacities due to air pollution. *Atmos. Environ.* **44**, 4606–4613 (2010).
- West, J. J., Fiore, A. M., Horowitz, L. W. & Mauzerall, D. L. Global health benefits of mitigating ozone pollution with methane emission controls. *Proc. Natl Acad. Sci. USA* **103**, 3988–3993 (2006).
- Duncan, B. N. *et al.* The influence of European pollution on ozone in the Near East and northern Africa. *Atmos. Chem. Phys.* **8**, 2267–2283 (2008).
- Liu, J., Mauzerall, D. L. & Horowitz, L. W. Evaluating inter-continental transport of fine aerosols: (2) Global health impact. *Atmos. Environ.* **43**, 4339–4347 (2009).
- Anenberg, S. C., Horowitz, L. W., Tong, D. Q. & West, J. J. An estimate of the global burden of anthropogenic ozone and fine particulate matter on premature human mortality using atmospheric modeling. *Environ. Health Perspect.* **118**, 1189–1195 (2010).
- Fann, N. *et al.* Estimating the national public health burden associated with exposure to ambient  $\text{PM}_{2.5}$  and ozone. *Risk Anal.* **32**, 81–95 (2012).
- Silva, R. A. *et al.* Global premature mortality due to anthropogenic outdoor air pollution and the contribution of past climate change. *Environ. Res. Lett.* **8**, <http://dx.doi.org/10.1088/1748-9326/8/3/034005> (2013).
- Lelieveld, J., Barlas, C., Giannadaki, D. & Pozzer, A. Model calculated global, regional and megacity premature mortality due to air pollution by ozone and fine particulate matter. *Atmos. Chem. Phys.* **13**, 7023–7037 (2013).

22. Giannadaki, D., Pozzer, A. & Lelieveld, J. Modeled global effects of airborne desert dust on air quality and premature mortality. *Atmos. Chem. Phys.* **14**, 957–968 (2014).
23. van Donkelaar, A. *et al.* Global estimates of ambient fine particulate matter concentrations from satellite-based aerosol optical depth: development and application. *Environ. Health Perspect.* **118**, 847–855 (2010).
24. Brauer, M. *et al.* Exposure assessment for estimation of the Global Burden of Disease attributable to outdoor air pollution. *Environ. Sci. Technol.* **46**, 652–660 (2012).
25. Thurston, G. D. *et al.* in *National Particle Component Toxicity (NPACT) Initiative: Integrated Epidemiologic and Toxicologic Studies of the Health Effects of Particulate Matter Components* (eds Lippmann, M. *et al.*) 127–166 (Health Effects Institute Research Report 177, Boston, 2013).
26. Lippmann, M., *et al.* (eds) *National Particle Component Toxicity (NPACT) Initiative: Integrated Epidemiologic and Toxicologic Studies of the Health Effects of Particulate Matter Components* (Health Effects Institute Research Report 177, Boston, 2013).
27. Vedal, S. *et al.* *National Particle Component Toxicity (NPACT) Initiative: Report on Cardiovascular Effects* (Health Effects Institute Research Report 178, Boston, 2013).
28. Yang, G. *et al.* Rapid health transition in China, 1990–2010: findings from the Global Burden of Disease Study 2010. *Lancet* **381**, 1987–2015 (2013).
29. Megaritis, A. G., Fountoukis, C., Charalampidis, P. E., Pilinis, C. & Pandis, S. N. Response of fine particulate matter concentrations to changes of emissions and temperature in Europe. *Atmos. Chem. Phys.* **13**, 3423–3443 (2013).
30. Ginoux, P., Prospero, J. M., Gill, T. E., Hsu, N. C. & Zhao, M. Global-scale attribution of anthropogenic and natural dust sources and their emission rates based on MODIS Deep Blue aerosol products. *Rev. Geophys.* **50**, RG3005 (2012).
31. Tager, I. *et al.* *Traffic-related Air Pollution: A Critical Review of the Literature on Emissions, Exposure, and Health Effects* (Health Effects Institute Special Report 17, Boston, 2010).
32. Pozzer, A. *et al.* Effects of business-as-usual anthropogenic emissions on air quality. *Atmos. Chem. Phys.* **12**, 6915–6937 (2012).

**Acknowledgements** We are grateful to the EDGAR team of the Joint Research Centre in Ispra, Italy, for the emission data. We acknowledge support from the Distinguished Scientist Fellowship Program at the King Saud University, Riyadh. The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013)/ERC grant agreement no. 226144.

**Author Contributions** J.L., A.P. and M.F. planned the research, A.P. performed the model calculations, J.L., A.P., D.G. and J.S.E. analysed the results, and J.L. and J.S.E. wrote the paper. All authors contributed to the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.L. ([jos.lelieveld@mpic.de](mailto:jos.lelieveld@mpic.de)).



## METHODS

**Model and emissions.** We used the global ECHAM5/MESSy atmospheric chemistry (EMAC)–general circulation model at a spatial resolution of T106L31, that is, with a spherical spectral truncation of T106, which corresponds to a quadratic Gaussian grid of approximately  $1.1^\circ \times 1.1^\circ$  latitude  $\times$  longitude ( $\sim 110$  km at the Equator), with 31 vertical hybrid terrain-following and pressure levels up to 10 hPa in the lower stratosphere. The core atmospheric model is the 5th generation European Centre Hamburg (ECHAM5, version 5.3.01) general circulation model<sup>33</sup>. EMAC includes sub-models that represent tropospheric and stratospheric processes and their interaction with oceans, land and human influences<sup>34–36</sup>. It uses the Modular Earth Submodel System (MESSy, v.1.09) to link submodels that describe emissions, atmospheric chemistry, aerosol and deposition processes; the results have been tested against *in situ* and remote sensing observations<sup>37–49</sup>.

Following up on Lelieveld *et al.*<sup>21</sup>, who focused on the year 2005, we present results for the years 2010, 2025 and 2050, applying monthly varying emission data from Doering *et al.*<sup>50</sup>, also used by Pozzer *et al.*<sup>32</sup>. The data are from the Emission Database for Global Atmospheric Research (EDGAR), prepared by the Joint Research Centre of the European Commission in Ispra (Italy) at a resolution of  $0.1^\circ$  latitude and longitude<sup>50,51</sup>. For the year 2010 we performed sensitivity calculations in which seven main emission categories have been removed one by one to compute the impact of these sources and to estimate their contributions to air quality control and related mortality. We first calculated the apportionment of source categories to the total  $\text{PM}_{2.5}$  and  $\text{O}_3$  concentrations and then applied the computed fractions to the total mortalities attributable to air pollution.

The categories are: (1) 'Natural' (NAT), mostly desert dust but locally also sea salt and dimethyl sulphide derived sulphate, some nitrate and ammonium from natural sources, volcanic sulphur emissions and organics released by the vegetation; (2) 'Industry' (IND), including iron and steel, chemical, pulp and paper, food, solvent and other manufacturing sectors, oil refineries and fuel production; (3) 'Land transport' (TRA), that is, road and non-road transport on land; (4) 'Residential and commercial energy use' (RCO), referring to local and commercial energy use from small combustion sources for space heating and cooking, including diesel generators and biofuel use; (5) 'Power generation' (PG), that is, public energy production by fossil fuel fired power plants; (6) 'Biomass burning' (BB), that is, tropical forest fires and deforestation, savanna and shrub fires, middle and high latitude forest and grassland fires, and agricultural waste burning; and (7) 'Agriculture' (AGR), dominated by ammonia emissions associated with the use of fertilizers and domesticated animals. Not included in these categories are air traffic and shipping. We find that the removal of individual source categories leads to a near-linear response in the modelled contributions to mortality, indicated by the small scaling corrections needed (about 10%) to add up to 100% in the country level contributions, that is, in Table 2 and Extended Data Table 3.

The BaU scenarios for 2025 and 2050 assume that energy and food consumption are largely determined by population growth and economic development, which in turn drive air pollution sources based on current legislation and technology<sup>32,50,51</sup>. This represents a pessimistic, but plausible future prospect. Comparable to Shindell *et al.*<sup>52</sup>, and different from the Representative Concentration Pathways of the Intergovernmental Panel on Climate Change<sup>53</sup>, the BaU scenario differentiates between air pollution and climate change mitigation measures, as the latter typically require relatively long-term and structural societal changes. The scenarios used here are based on projections for energy and fuel computed by the Prospective Outlook for the Long-term Energy System (POLES) model<sup>51,54</sup> and for agriculture, land-use and waste projections by the Integrated Model to Assess the Global Environment (IMAGE)<sup>55</sup>.

The population development in the BaU scenario is consistent with our mortality calculations, as described below, projecting 9 billion people in 2050. For additional details we refer to Pozzer *et al.*<sup>32</sup> and references therein. While BaU projections should not be conceived as 'predictions', especially for 2050, they represent the current trajectory into the future and may be considered a worst-case scenario, to explore what can be expected if air quality policies and health care remain as they are today. Note that these results are not sensitive to differential toxicity assumptions as the total mortality induced by  $\text{PM}_{2.5}$  is not affected, only the attribution to source categories. For the future scenarios we used the baseline mortalities for 2010. Hence the implicit assumption is that smoking habits, diets and health care remain unchanged.

The model meteorology has been forced by pre-calculated sea surface temperatures and ice coverage based on a 10-year climatology (2000–2009) adopted from the AMIP-II database<sup>56,57</sup>. The model was applied in atmospheric chemistry–transport mode by switching the coupling between radiation and atmospheric chemistry off, so that atmospheric composition changes do not influence the model dynamics<sup>32</sup>. This is justified considering that air quality projections are primarily driven by emissions rather than climate change<sup>58,59</sup>, even though natural sources, biomass burning and deposition processes can be influenced by climatic

conditions<sup>20,59–62</sup>. For example, Fang *et al.*<sup>62</sup> project a 4% climate change effect for  $\text{PM}_{2.5}$  related mortality and less than 1% for  $\text{O}_3$  related mortality by the end of the 21st century.

Although our model resolution does not resolve small-scale heterogeneities in the urban environment, a comparison with satellite and ground-based remote sensing observations indicates that this is not critical. The exposure response functions used to calculate mortalities are based on annual mean concentrations for which these heterogeneities largely average out. This is illustrated by Extended Data Fig. 3, which compares a simulation for the year 2010 with ground-based AERONET remote sensing data of aerosol optical depth (AOD) (<http://aeronet.gsfc.nasa.gov>). Since our model approximates though not replicates meteorological conditions for the year 2010, and local flows near the AERONET stations cannot be captured, substantial scatter around the ideal 1:1 comparison is expected. The comparison shows that the model mean error and bias are small (the latter absent for the annual mean), and the correlation good. We have also performed a comparison between MODIS (satellite) and AERONET data of AOD, leading to similar spread and correlations, the latter also increasing through averaging (not shown).

The primary differences in the relationships between emissions and exposures for ground level sources, such as traffic, in comparison with elevated sources, such as power plants, have been accounted for in our model<sup>43</sup>. The relative impacts of secondary particles (such as sulfates and nitrates) from these sources are expected to be realistically simulated. On the other hand, models such as ours cannot capture the fine structure of near-source gradients in ultrafine PM along transportation corridors. Because of this our estimates of the relative impacts of urban traffic and urban sources of primary fine particles may be biased downward, though only to the extent that ultrafine PM is in fact responsible for the mortality seen in cohort studies. As discussed above, the relative toxicity of various constituents of ambient  $\text{PM}_{2.5}$  has not been well established. Our sense is that the sensitivity study, allowing for carbonaceous particles to be five times as toxic as sulfates, nitrates and crustal material, is adequate to cover any potential differences in the relationships between emissions, exposure and differential toxicity of traffic related  $\text{PM}_{2.5}$ .

To investigate if our model reproduces urban concentration increments of  $\text{PM}_{2.5}$  and  $\text{O}_3$ , that is, comparing the urban background with the rural environment, we compare our results with recent case studies<sup>63–67</sup>. For Paris and London our model computes urban  $\text{PM}_{2.5}$  increments of 18% and 2%, respectively, consistent with the measurements and highly resolved model calculations. Our model calculations suggest that the leading sources of  $\text{PM}_{2.5}$  in Paris are residential energy use, agriculture and traffic. Agricultural emissions ( $\text{NH}_3/\text{NH}_4^+$ ) are transported from the rural environment and contribute to  $\text{PM}_{2.5}$  in the city. For London we calculate that  $\text{PM}_{2.5}$  is most strongly influenced by agriculture, traffic and power generation. The limited contribution by land traffic and the importance of atmospheric transport for air quality in London have been corroborated by observational analysis<sup>63</sup>. For Beijing we calculate an urban  $\text{PM}_{2.5}$  increment of 5%, consistent with the conclusion by Zhang *et al.*<sup>67</sup> that regional sources are crucial contributors to  $\text{PM}_{2.5}$ . They estimate the contribution by traffic and waste incineration at 4%; our results suggest that traffic alone contributes 3% in this city and residential energy use 47%, which we find to be representative of China (Table 2).

Our model calculations indicate that these relatively small urban increments for  $\text{PM}_{2.5}$  are typical for many, though not all, cities. For example, for Johannesburg (including Pretoria) we find +41% and for the Pearl River area +62%, and in both conurbations residential energy use is the leading source of  $\text{PM}_{2.5}$ . For  $\text{O}_3$  we find generally small and negative urban increments due to titration of  $\text{O}_3$  by local traffic emissions (in Paris –7% and in London –5%). Negative urban increments due to NO by traffic of a few per cent (comparing weekend with weekdays) have also been documented for American cities<sup>68</sup>. For Chicago, New York, Los Angeles and Atlanta we find negative  $\text{O}_3$  increments of 1–5% due to traffic and power generation.

**Sample size.** No statistical methods were used to predetermine sample size.

**Exposure response functions.** The premature mortality attributable to  $\text{PM}_{2.5}$  and  $\text{O}_3$  has been calculated by applying the EMAC model for the present (2010) and projected future (2025, 2050) concentrations. We combined the results with epidemiological exposure response functions by employing the following relationship to estimate the excess (that is, premature) mortality:

$$\Delta\text{Mort} = y_o[(\text{RR} - 1)/\text{RR}]\text{Pop} \quad (1)$$

$\Delta\text{Mort}$  is a function of the baseline mortality rate due to a particular disease category  $y_o$  for countries and/or regions estimated by the World Health Organization<sup>69</sup> (the regions and strata are listed in the Extended Data Table 1). The term  $(\text{RR} - 1)/\text{RR}$  is the attributable fraction and RR is the relative risk. The disease specific baseline mortality rates have been obtained from the WHO Health Statistics and Health Information System. The value of RR is calculated for the different disease categories attributed to  $\text{PM}_{2.5}$  and  $\text{O}_3$  for the population below 5 years of age (ALRI) and 30 years and older (IHD, CEV, COPD, LC) using

exposure response functions from the 2010 GBD analysis of the WHO (and described below).

The population (Pop) data for regions, countries and urban areas have been obtained from the NASA Socioeconomic Data and Applications Center (SEDAC), hosted by the Columbia University Center for International Earth Science Information Network (CIESIN), available at a resolution of  $2.5' \times 2.5'$  (about  $5 \text{ km} \times 5 \text{ km}$ ) (<http://sedac.ciesin.columbia.edu/>), and projections by the United Nations Department of Economic and Social Affairs/Population Division<sup>70</sup> (<http://esa.un.org/unpd/wpp/>). Urban areas are defined by applying a population density threshold of 400 individuals per  $\text{km}^2$ , while for megacities and major conurbations the threshold is 2,000 individuals per  $\text{km}^2$ . We note that the resolution of our atmospheric model, about  $1^\circ$  latitude/longitude, is coarser than that of the population data, and our model does not resolve details of the urban environment. However, our anthropogenic emission data are aggregated from a resolution of 10 km to that of the model grid, accounting for relevant details such as altitude dependence (for example, stack emissions and hot plume rise effects)<sup>43</sup>.

Lelieveld *et al.*<sup>21</sup> (henceforth L2013) derived the relative risk RR from the following exposure response function:

$$\text{RR} = \exp[b(X - X_o)] \quad (2)$$

The term  $X$  represents the model calculated annual mean concentration of  $\text{PM}_{2.5}$  or  $\text{O}_3$ . The value of  $X_o$  is the threshold concentration below which no additional risk is assumed (concentration–response threshold). The parameter  $b$  is the concentration response coefficient. However, it has been argued that this expression is based on epidemiological cohort studies in the USA and Europe where annual mean  $\text{PM}_{2.5}$  concentrations are typically below  $30 \mu\text{g m}^{-3}$ , which may not be representative for countries where air pollution levels can be much higher, for example in South and East Asia. This is particularly relevant for our BaU scenario. Therefore, here we have used the revised exposure response function of Burnett *et al.*<sup>8</sup> who also included epidemiological data from the exposure to second-hand smoke, indoor air pollution and active smoking to account for high  $\text{PM}_{2.5}$  concentrations, and tested eight different expressions. The best fit to the data was found for the following relationship, which was also used by Lim *et al.*<sup>5</sup> for the GBD for the year 2010:

$$\text{RR} = 1 + a\{1 - \exp[-b(X - X_o)^p]\} \quad (3)$$

The RR functions were derived by Burnett *et al.*<sup>8</sup>. We applied this model for the different categories, represented by their figures 1 and 2, shown to be superior to other forms previously used in burden assessments. We also adopted the upper and lower bounds, likewise shown in these figures, representing the 95% confidence intervals (CI95). The latter were derived based on Monte Carlo simulations, leading to 1,000 sets of coefficients and exposure response functions from which the upper and lower bounds were calculated.

Following Burnett *et al.*<sup>8</sup> and Lim *et al.*<sup>5</sup> we combine all aerosol types, hence including natural particulates such as desert dust. Note that by using  $\text{PM}_{2.5}$  mass, we do not distinguish the possibly different toxicity of various kinds of particles. This information is not available from epidemiological cohort studies, but could potentially substantially affect both our overall estimates of mortality and the geographical patterns. This is addressed by sensitivity calculations presented in the main text, Table 2 and Extended Data Fig. 1. For COPD related to  $\text{O}_3$  we applied the exposure response function by Ostro *et al.*<sup>3</sup>:

$$\text{RR} = [(X + 1)/(X_o + 1)]^b \quad (4)$$

where  $b$  is 0.1521 and  $X_o$  the average of the range 33.3–41.9 p.p.b.v.  $\text{O}_3$  indicated by Lim *et al.*<sup>5</sup>, that is, 37.6 p.p.b.v. Previously we used model calculated pre-industrial  $\text{O}_3$  concentrations to estimate  $X - X_o$  (ref. 21), leading to about 20% higher estimates for mortality by ‘respiratory disease’ related solely to  $\text{O}_3$  compared to the current estimate for COPD due to both  $\text{PM}_{2.5}$  and  $\text{O}_3$ .

For detailed discussion of uncertainties and sensitivity calculations that address the shape of exposure response functions, we refer to earlier work<sup>5,8,21,22</sup> and references therein. L2013 estimated statistical uncertainties by propagating the quantified (random) errors of all parameters in the exposure response functions. They found that the CI95 of estimated mortality attributable to air pollution in Europe, North and South America, South and East Asia are within 40%, whereas they are 100–170% in Africa and the Middle East. Our results are very close to the GBD, which substantiates the estimates by Lim *et al.*<sup>5</sup> and provides consistency with the most recent estimates for 2010, serving as a basis for our investigations.

We emphasize that the confidence intervals described here, and those reported by Lim *et al.*<sup>5</sup>, reflect only the statistical uncertainty of the parameters used in the concentration–response functions. It is known that the uncertainty in interpretation of epidemiological results can be dominated by other model or epistemic uncertainties, such as those having to do with the control of confounders. Sources of uncertainty have been summarized by Kinney *et al.*<sup>71</sup>, who underscore the need to determine the differential toxicity of specific component species within the

complex mixture of particulate matter. Our sensitivity calculations (Table 2 and Extended Data Fig. 1) corroborate that this can have significant influence, especially in areas where carbonaceous compounds contribute strongly to  $\text{PM}_{2.5}$ .

We emphasize the dearth of studies that link  $\text{PM}_{2.5}$  from biomass combustion emissions—rich in carbonaceous particles—to IHD. Expert judgment studies on the toxicity of particulate matter have reported uncertainties much larger than those suggested by analysis of parameter uncertainty alone<sup>10,72</sup>. Although the CI95 intervals provided above include a larger range of parameters and uncertainties than these earlier studies, they should be viewed as lower bounds on the true uncertainty in estimates of the health effects of  $\text{PM}_{2.5}$  exposure, especially  $\text{PM}_{2.5}$  from biomass burning and biofuel use. If we consider the possibility that biomass burning (BB, including agricultural waste burning) and residential energy use (RCO, dominated by biofuel use) do not contribute to mortality by IHD, the total mortality attributable to air pollution would decrease from 3.3 to 3.0 million per year (Extended Data Table 7). The largest effect is found in Southeast Asia where biomass combustion (RCO and BB) is a main source of air pollution. While the global contribution by residential energy use, as presented in Table 2, would decrease from 31% to 26%, and of biomass burning from 5% to 4% (the other categories increase proportionally), the ranking of the different sources and hence our conclusions remain unchanged, as RCO and BB would still be the largest and smallest source category, respectively.

Issues such as the shape of the concentration–response functions and the existence and specific levels of concentration–response thresholds have been discussed by the experts<sup>10,71,72</sup>. These have been accounted for by Burnett *et al.*<sup>8</sup>, however, uncertainty related to the differences in central estimates given by various cohort studies is not reflected in the estimates of parameter uncertainty by Lim *et al.*<sup>5</sup>. This problem has grown more substantial recently as the results from new cohort studies have become available<sup>73</sup>. Furthermore, uncertainty about the relative toxicity of different constituents of  $\text{PM}_{2.5}$  remains. Since the current study underscores that the sources of mortality attributable to  $\text{PM}_{2.5}$  can differ strongly between different regions (Fig. 2), this aspect merits greater attention in future.

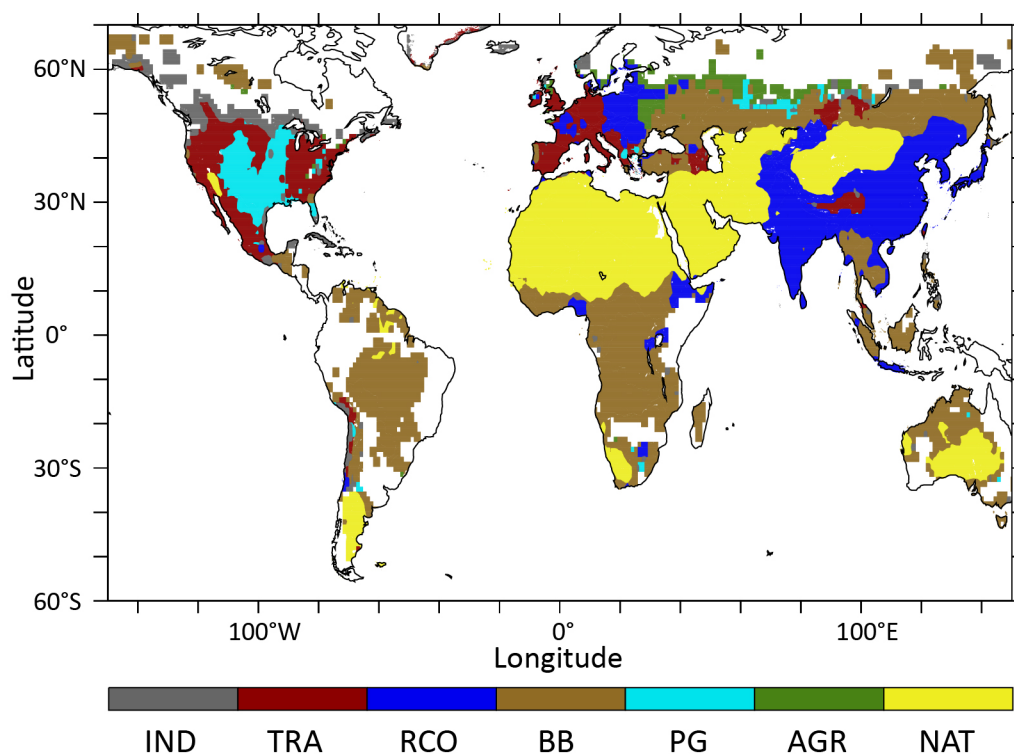
**Comparison to previous work.** We estimate the combined ( $\text{PM}_{2.5}$  and  $\text{O}_3$  related) global mortality attributable to air pollution in 2010 at 3.3 million. Our global estimate for  $\text{PM}_{2.5}$  related mortality of 3.15 million per year is close to that of 3.22 million per year in the GBD study for 2010 (ref. 4). However, it is substantially higher than the recent multi-model study of Silva *et al.*<sup>20</sup> for the year 2000, being 2.1 million per year. The difference can be explained by the focus of Silva *et al.*<sup>20</sup> on anthropogenic pollution in 2000, whereas our study and the GBD account for emission increases between 2000 and 2010 and also include natural sources.

Our global estimate of  $\text{O}_3$  related mortality by COPD in 2010 is 142,000, substantially lower than the estimates of Anenberg *et al.*<sup>18</sup>, 700,000 deaths in 2000; L2013, 773,000 in 2005; and Silva *et al.*<sup>20</sup>, 470,000 deaths in 2000; but quite close to the GBD estimate of 152,000 deaths in 2010. Much of the difference between our results (and those from the 2010 GBD) and previous work is explained by the fact that we attribute COPD to both  $\text{O}_3$  and  $\text{PM}_{2.5}$ . When our results for COPD from both  $\text{O}_3$  and  $\text{PM}_{2.5}$  are combined, our overall estimate of COPD mortality from air pollution agrees with the above-mentioned studies within about 25–30%. The remaining differences are largely due to the use of a concentration response threshold,  $X_o$ , in our new work, which substantially reduces mortality estimates. Anenberg *et al.*<sup>18</sup> and L2013 did not apply a threshold but computed the natural background based on preindustrial emissions. In these analyses the calculated ambient concentrations are typically lower than  $X_o$ . For example, the global average  $\text{O}_3$  ambient concentration at the surface in our pre-industrial simulation is 19 p.p.b.v. The global mortality estimate for 2010 presented here is 10% higher than that of L2013 for 2005. This is primarily due to the fact that we also account for natural sources in the present work. If we subtract the natural fraction, our estimate of mortality attributable to anthropogenic air pollution for 2010 is 9% lower than that of L2013, mostly related to the new exposure response functions applied here.

Our calculations suggest that natural sources contribute relatively strongly to mortality attributable to air pollution (18%), about 600,000 per year, which is to a large degree caused by airborne desert dust. Recently we reported a global dust-related mortality rate of about 400,000 per year, substantially lower than the present estimate<sup>22</sup>. While here we follow the GBD methodology<sup>5</sup>, it is likely to yield an upper limit. Instead of the annual mean dust concentrations Giannadaki *et al.*<sup>22</sup> used the median concentrations, motivated by the intermittent nature of dust events. Their sensitivity calculations indicate that had they used the mean concentration instead, their estimate of global dust-related mortality would have increased from 402,000 per year to 622,000 per year. Finally, if we assume that carbonaceous aerosols are five times more toxic than other compounds, including dust particles, the contribution by natural sources would decrease from about 600,000 per year (18%) to 360,000 per year (11%).

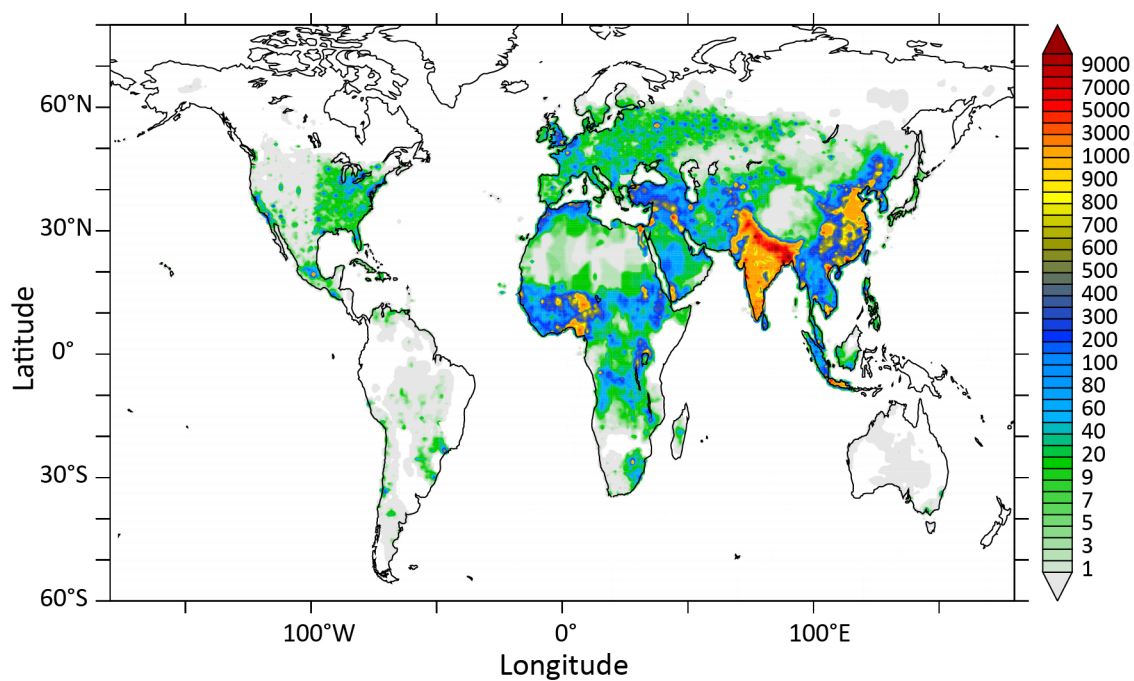
33. Roeckner, E. *et al.* Sensitivity of simulated climate to horizontal and vertical resolution in the ECHAM5 atmosphere model. *J. Clim.* **19**, 3771–3791 (2006).
34. Jöckel, P. *et al.* Technical Note: The Modular Earth Submodel System (MESSy) – a new approach towards earth system modeling. *Atmos. Chem. Phys.* **5**, 433–444 (2005).
35. Jöckel, P. *et al.* The atmospheric chemistry general circulation model ECHAM5/MESSy: Consistent simulation of ozone from the surface to the mesosphere. *Atmos. Chem. Phys.* **6**, 5067–5104 (2006).
36. Pozzer, A., Jöckel, P., Kern, B. & Haak, H. The atmosphere-ocean general circulation model EMAC-MPIOM. *Geosci. Model Dev.* **4**, 771–784 (2011).
37. Sander, R., Kerkweg, A., Jöckel, P. & Lelieveld, J. Technical note: The new comprehensive atmospheric chemistry module MECCA. *Atmos. Chem. Phys.* **5**, 445–450 (2005).
38. Kerkweg, A. *et al.* Technical Note: An implementation of the dry removal processes DRY DEPosition and SEDImentation in the Modular Earth Submodel System (MESSy). *Atmos. Chem. Phys.* **6**, 4617–4632 (2006).
39. Tost, H. *et al.* Technical note: A new comprehensive SCAVenging submodel for global atmospheric chemistry modeling. *Atmos. Chem. Phys.* **6**, 565–574 (2006).
40. Tost, H. *et al.* Global cloud and precipitation chemistry and wet deposition: tropospheric model simulations with ECHAM5/MESSy1. *Atmos. Chem. Phys.* **7**, 2733–2757 (2007).
41. Pozzer, A. *et al.* Technical Note: The MESSy-submodel AIRSEA calculating the air-sea exchange of chemical species. *Atmos. Chem. Phys.* **6**, 5435–5444 (2006).
42. Pozzer, A. *et al.* Simulating organic species with the global atmospheric chemistry general circulation model ECHAM5/MESSy1: a comparison of model results with observations. *Atmos. Chem. Phys.* **7**, 2527–2550 (2007).
43. Pozzer, A., Jöckel, P. & van Aardenne, J. The influence of the vertical distribution of emissions on tropospheric chemistry. *Atmos. Chem. Phys.* **9**, 9417–9432 (2009).
44. Pozzer, A. *et al.* Distributions and regional budgets of aerosols and their precursors simulated with the EMAC chemistry-climate model. *Atmos. Chem. Phys.* **12**, 961–987 (2012).
45. Astitha, M. *et al.* Parameterization of dust emissions in the global atmospheric chemistry-climate model EMAC: impact of nudging and soil properties. *Atmos. Chem. Phys.* **12**, 11057–11083 (2012).
46. Pringle, K. J. *et al.* Description and evaluation of GMXe: A new aerosol submodel for global simulations (v1). *Geosci. Model Dev.* **3**, 391–412 (2010).
47. Pringle, K. J. *et al.* Global distribution of the effective aerosol hygroscopicity parameter for CCN activation. *Atmos. Chem. Phys.* **10**, 5241–5255 (2010).
48. de Meij, A. *et al.* EMAC model evaluation and analysis of atmospheric aerosol properties and distribution. *Atmos. Res.* **114–115**, 38–69 (2012).
49. Christoudias, T. & Lelieveld, J. Modelling the global atmospheric transport and deposition of radionuclides from the Fukushima Dai-ichi nuclear accident. *Atmos. Chem. Phys.* **13**, 1425–1438 (2013).
50. Doering, U., Janssens-Maenhout, G., van Aardenne, J. & Pagliari, V. *Climate Change and Impact Research in the Mediterranean Environment: Scenarios of Future Climate Change*. JRC Tech. Note 62957 (Joint Research Centre, Ispra, 2010).
51. Van Aardenne, J. *et al.* *Climate and Air Quality Impacts of Combined Climate Change and Air Pollution Policy Scenarios*. JRC Sci. Tech. Rep. 61281 <http://dx.doi.org/10.2788/33719> (Joint Research Centre, Ispra, 2010).
52. Shindell, D. *et al.* Simultaneously mitigating near-term climate change and improving human health and food security. *Science* **335**, 183–189 (2012).
53. Stocker, T. F. *et al.* (eds) *Climate Change 2013: The Physical Science Basis* (Cambridge Univ. Press, 2013).
54. Russ, P., Wiesenthal, T., van Regenmortel, D. & Ciscar, J. C. *Global Climate Policy Scenarios for 2030 and Beyond. Analysis of Greenhouse Gas Emission Reduction Pathway Scenarios with the POLES and GEM-E3 models*. JRC Ref. Rep. EUR 23032 EN, <http://ipts.jrc.ec.europa.eu/publications/pub.cfm?id=1510> (Joint Research Centre, Ispra, 2007).
55. Bouwman, A. F., Kram, T. & Klein Goldewijk, K. (eds) *Integrated Modelling of Global Environmental change. An Overview of IMAGE 2.4* (Netherlands Environmental Assessment Agency (MNP), Bilthoven, 2006).
56. Taylor, K., Williamson, D. & Zwiers, F. *The Sea Surface Temperature and Sea Ice Concentration Boundary Conditions for AMIP II Simulations*. PCMDI Tech. Rep. 60 (Program for Climate Model Diagnosis and Intercomparison, Lawrence Livermore National Laboratory, Livermore, California, 2000).
57. Hurrell, J. *et al.* A new sea surface temperature and sea ice boundary dataset for the Community Atmosphere Model. *J. Clim.* **21**, 5145–5153 (2008).
58. Jacob, D. J. & Winner, D. A. Effect of climate change on air quality. *Atmos. Environ.* **43**, 51–63 (2009).
59. Pye, H. O. T. *et al.* Effect of changes in climate and emissions on future sulfate-nitrate-ammonium aerosol levels in the United States. *J. Geophys. Res.* **114**, D01205, <http://dx.doi.org/10.1029/2008JD010701> (2009).
60. Hedegaard, G. B., Christensen, J. H. & Brandt, J. The relative importance of impacts from climate change vs. emissions change on air pollution levels in the 21st century. *Atmos. Chem. Phys.* **13**, 3569–3585 (2013).
61. Naik, V. *et al.* Preindustrial to present-day changes in tropospheric hydroxyl radical and methane lifetime from the Atmospheric Chemistry and Climate Model Intercomparison Project (ACCMIP). *Atmos. Chem. Phys.* **13**, 5277–5298 (2013).
62. Fang, Y. *et al.* Impacts of 21st century climate change on global air pollution-related premature mortality. *Clim. Change* **121**, 239–253 (2013).
63. Jones, A. M., Yin, J. & Harrison, R. M. The weekday-weekend difference and the estimation of the non-vehicle contributions to the urban increment of airborne particulate matter. *Atmos. Environ.* **42**, 4467–4479 (2008).
64. Harrison, R. M., Laxen, D., Moorcroft, S. & Laxen, K. Processes affecting concentrations of fine particulate matter (PM<sub>2.5</sub>) in the UK atmosphere. *Atmos. Environ.* **46**, 115–124 (2012).
65. Moussiopoulos, N. *et al.* An approach for determining urban concentration increments. *Int. J. Environ. Pollut.* **50**, 376–385 (2012).
66. Timmermans, R. M. A. *et al.* Quantification of the urban air pollution increment and its dependency on the use of down-scaled and bottom-up city emission inventories. *Urban Clim.* **6**, 44–62 (2013).
67. Zhang, R. *et al.* Chemical characterization and source apportionment of PM<sub>2.5</sub> in Beijing: seasonal perspective. *Atmos. Chem. Phys.* **13**, 7053–7074 (2013); *Atmos. Chem. Phys.* **14**, 175 (2014).
68. Blanchard, C. L., Tanenbaum, S. & Lawson, D. R. Differences between weekday and weekend air pollutant levels in Atlanta; Baltimore; Chicago; Dallas–Fort Worth; Denver; Houston; New York; Phoenix; Washington, DC; and surrounding areas. *J. Air Waste Manag. Assoc.* **58**, 1598–1615 (2008).
69. World Health Organization. *World Health Organization Statistical Information System (WHOSIS), Detailed Data Files of the WHO Mortality Database* [http://www.who.int/healthinfo/statistics/mortality\\_rawdata/en/](http://www.who.int/healthinfo/statistics/mortality_rawdata/en/) (WHO, Geneva, 2012).
70. United Nations Department of Economic and Social Affairs/Population Division. *World Population Prospects: the 2004 Revision*. E.05.XIII.12 (United Nations, 2005).
71. Kinney, P. L. *et al.* On the use of expert judgment to characterize uncertainties in the health benefits of regulatory controls of particulate matter. *Environ. Sci. Policy* **13**, 434–443 (2010).
72. Roman, H. A. *et al.* Expert judgment assessment of the mortality impact of changes in ambient fine particulate matter in the U.S. *Environ. Sci. Technol.* **42**, 2268–2274 (2008).
73. Cao, J. *et al.* Association between long-term exposure to outdoor air pollution and mortality in China: A cohort study. *J. Hazard. Mater.* **186**, 1594–1600 (2011).



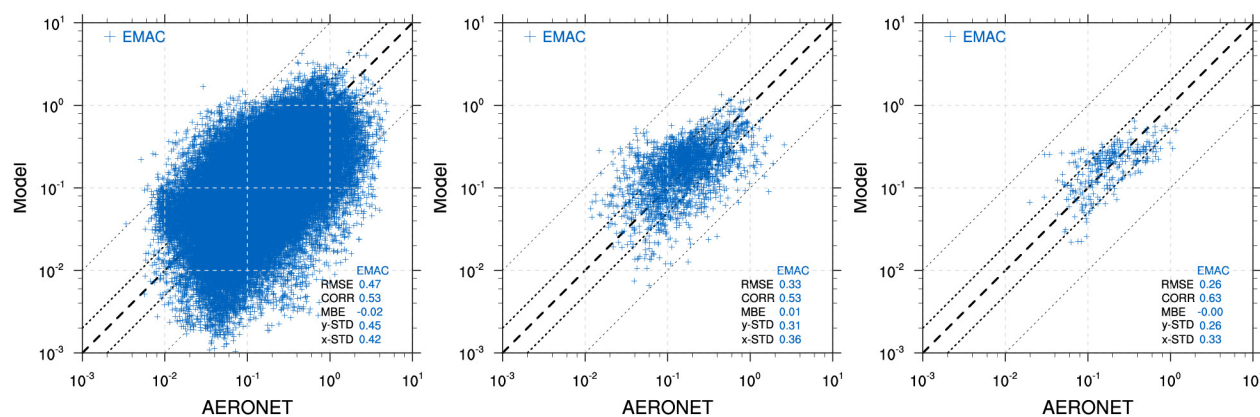


**Extended Data Figure 1 | Source categories responsible for the largest impact on mortality linked to outdoor air pollution in 2010 from a sensitivity calculation with carbonaceous aerosol having a five times larger**

**impact than inorganic and crustal compounds.** IND, industry; TRA, land traffic; RCO, residential energy use (for example, heating, cooking); BB, biomass burning; PG, power generation; AGR, agriculture; and NAT, natural.



**Extended Data Figure 2 | Increase in mortality linked to outdoor air pollution from 2010 to 2050 (business-as-usual scenario).** Units (colour coded), deaths per area of 100 km × 100 km. In the white areas, no additional mortality is projected.



**Extended Data Figure 3 | Comparison of EMAC model calculated aerosol optical depth (AOD) with AERONET observations, using all available measurements worldwide in the year 2010.** Although the comparison with individual data points shows a large scatter (left panel), the bias is small (MBE), and time averaging improves the agreement. The middle panel shows a comparison of the monthly means, and the right panel the annual

means (that is, showing individual stations) for which the mean error (root mean square error, RMSE) is smallest, the correlation highest and the bias absent. The long-dashed line indicates absolute agreement, the bold short-dashed lines agreement within a factor of two and the short-dashed lines agreement within a factor of ten.



**Extended Data Table 1 | WHO regions, mortality strata, child and adult mortality characteristics, and the countries and territories included**

Region	Stratum	Child mortality	Adult mortality	Countries and territories within stratum
Africa	Afr-D	High	High	Algeria, Angola, Benin, Burkina Faso, Cameroon, Cape Verde, Chad, Comoros, Equatorial Guinea, Gabon, Gambia, Ghana, Guinea, Guinea-Bissau, Liberia, Madagascar, Mali, Mauritania, Mauritius, Mayotte, Niger, Nigeria, Reunion, Saint Helena, Sao Tome and Principe, Senegal, Seychelles, Sierra Leone, Togo
	Afr-E	High	Very high	Botswana, Burundi, Central African Republic, Congo, Côte d'Ivoire, Democratic Republic of the Congo, Eritrea, Ethiopia, Kenya, Lesotho, Malawi, Mozambique, Namibia, Rwanda, South Africa, Swaziland, Uganda, United Republic of Tanzania, Zambia, Zimbabwe
Americas	Amr-A	Very low	Very low	Canada, Cuba, Greenland, Saint Pierre and Miquelon, United States of America
	Amr-B	Low	Low	Anguilla, Antigua and Barbuda, Argentina, Aruba, Bahamas, Barbados, Belize, Bermuda, Brazil, British Virgin Islands, Cayman Islands, Chile, Colombia, Costa Rica, Dominica, Dominican Republic, El Salvador, Falkland Islands, French Guiana, Grenada, Guadeloupe, Guyana, Honduras, Jamaica, Martinique, Mexico, Montserrat, Netherlands Antilles, Panama, Paraguay, Puerto Rico, Saint Kitts and Nevis, Saint Lucia, Saint Vincent and the Grenadines, Suriname, Trinidad and Tobago, Turks and Caicos Islands, United States Virgin Islands, Uruguay, Bolivarian Republic of Venezuela
	Amr-D	High	High	Bolivia, Ecuador, Guatemala, Haiti, Nicaragua, Peru
	Sear-B	Low	Low	Indonesia, Sri Lanka, Thailand
Southeast Asia	Sear-D	High	High	Bangladesh, Bhutan, Democratic People's Republic of Korea, East Timor, India, Maldives, Myanmar, Nepal
	Eur-A	Very low	Very low	Andorra, Austria, Belgium, Croatia, Cyprus, Czech Republic, Denmark, Faeroe Islands, Finland, France, Germany, Gibraltar, Greece, Guernsey, Iceland, Ireland, Isle of Man, Israel, Italy, Jersey, Liechtenstein, Luxembourg, Malta, Monaco, Netherlands, Norway, Portugal, San Marino, Slovenia, Spain, Svalbard, Sweden, Switzerland, United Kingdom
	Eur-B	Low	Low	Albania, Armenia, Azerbaijan, Bosnia and Herzegovina, Bulgaria, Georgia, Kyrgyzstan, Poland, Romania, Serbia and Montenegro, Slovakia, Tajikistan, The former Yugoslav Republic of Macedonia, Turkey, Turkmenistan, Uzbekistan
Europe	Eur-C	Low	High	Belarus, Estonia, Hungary, Kazakhstan, Latvia, Lithuania, Republic of Moldova, Russia, Ukraine
	Emr-B	Low	Low	Bahrain, Iran, Jordan, Kuwait, Lebanon, Libyan Arab Jamahiriya, Oman, Qatar, Saudi Arabia, Syrian Arab Republic, Tunisia, United Arab Emirates
	Emr-D	High	High	Afghanistan, Djibouti, Egypt, Iraq, Morocco, Palestinian Territories, Pakistan, Somalia, Sudan, Yemen
Eastern Mediterranean	Wpr-A	Very low	Very low	Australia, Brunei Darussalam, Japan, New Zealand, Singapore
	Wpr-B	Low	Low	Cambodia, China, Cook Islands, Fiji, French Polynesia, Guam, Hong Kong, Kiribati, Lao People's Democratic Republic, Macao, Malaysia, Marshall Islands, Pitcairn, Fed. States of Micronesia, Mongolia, Nauru, New Caledonia, Niue, Norfolk Island, Northern Mariana Islands, Palau, Papua New Guinea, Philippines, Republic of Korea, Samoa, Solomon Islands, Taiwan, Tokelau, Tonga, Tuvalu, Vanuatu, Vietnam, Wallis and Futuna
Western Pacific				

Extended Data Table 2 | Premature mortality related to PM<sub>2.5</sub> and O<sub>3</sub> in 2010

Mortality attributable to air pollution (deaths ×10 <sup>3</sup> )									
		PM <sub>2.5</sub>					O <sub>3</sub>	Total	
	Strata	Population (×10 <sup>6</sup> )	ALRI <5 yr	IHD ≥30 yr	CEV ≥30 yr	COPD ≥30 yr	LC ≥30 yr	COPD ≥30 yr	
Africa	Afr-D	379	77	37	62	9	1	2	188
	Afr-E	430	13	17	15	3	0	1	49
Americas	Amr-A	352	0	38	6	4	6	3	57
	Amr-B	493	0	5	1	0	1	1	8
	Amr-D	85	0	0	0	0	0	0	1
Eastern Mediterranean	Emr-B	165	2	34	20	2	1	3	62
	Emr-D	437	54	80	67	11	3	9	224
Europe	Eur-A	410	0	73	33	7	16	3	132
	Eur-B	229	1	57	34	3	7	2	104
	Eur-C	228	0	110	28	2	4	0	144
Southeast Asia	Sear-B	324	3	37	23	7	3	2	75
	Sear-D	1,438	61	290	227	117	12	80	787
Western Pacific	Wpr-A	156	0	11	9	0	5	2	27
	Wpr-B	1,656	19	288	784	209	102	33	1,435
World		6,783	230	1,079	1,311	374	161	142	3,297

**Extended Data Table 3 | Premature mortality by PM<sub>2.5</sub> and O<sub>3</sub> related diseases in 2010 in countries where it exceeds 9,000 individuals per year (<5 and ≥30 years old)**

Country	Deaths in 2010	Natural	Industry	Land traffic	Residential energy	Power generation	Biomass burning	Agriculture
China	1,357,353	118,954	106,754	44,751	435,763	237,324	18,414	395,390
India	644,993	74,145	42,336	30,070	325,604	89,130	42,163	41,541
Pakistan	110,571	63,147	2,478	3,389	34,707	2,761	2,108	1,977
Bangladesh	91,923	0	6,117	5,656	50,382	13,697	6,418	9,652
Nigeria	89,022	68,479	176	85	12,006	258	7,554	462
Russia	67,152	630	5,193	7,731	4,885	14,606	5,477	28,628
USA	54,905	1,290	3,297	11,435	3,192	16,929	2,537	16,221
Indonesia	52,417	71	1,814	1,244	31,498	2,379	14,338	1,070
Ukraine	51,238	55	4,632	5,188	3,011	9,459	2,326	26,563
Vietnam	44,097	0	3,627	1,686	22,575	5,486	5,378	5,343
Egypt	35,322	32,651	210	450	190	816	61	941
Germany	34,422	0	4,452	6,928	2,684	4,402	279	15,675
Turkey	31,943	4,912	3,414	3,487	2,812	6,194	1,851	9,269
Iran	26,108	21,175	662	969	311	1,101	230	1,656
Japan	25,516	0	4,567	2,526	3,046	4,458	1,154	9,763
Sudan	24,255	22,249	59	47	200	133	1,488	77
Myanmar	22,537	10	1,082	842	8,287	2,662	8,707	944
Italy	20,809	1,251	2,930	3,519	1,454	3,192	376	8,085
Iraq	20,335	18,513	209	390	109	510	91	510
Thailand	19,843	0	2,211	1,469	5,207	2,944	6,529	1,481
France	17,800	0	2,515	3,152	2,468	2,113	211	7,339
Dem. Rep. Korea (N)	16,783	0	1,996	770	3,445	3,467	715	6,386
United Kingdom	15,488	0	1,627	3,091	854	2,412	63	7,438
Algeria	14,954	11,262	1,113	656	194	773	107	847
Dem. Republic Congo	14,880	901	193	45	1,405	121	12,119	92
Romania	14,633	0	1,336	1,825	1,403	3,225	573	6,270
Saudi Arabia	14,600	13,708	165	165	46	308	38	167
Poland	14,561	0	1,451	1,886	2,273	2,265	372	6,310
Korea (S)	14,352	0	2,045	1,108	2,049	2,803	321	6,024
Morocco (+W. Sahara)	14,217	10,929	966	346	224	856	99	795
Niger	13,061	12,893	9	0	89	16	32	19
Uzbekistan	11,598	7,341	671	590	578	623	265	1,526
Nepal	10,926	56	641	510	7,481	1,090	681	465
Mali	9,444	9,060	1	0	101	0	273	6
Ghana	9,317	5,552	105	68	525	68	2,922	75
Burkina Faso	9,295	8,851	3	0	127	20	256	35
World	3,297,370	596,895	226,137	163,852	1,002,370	464,748	179,268	664,100

Extended Data Table 4 | Premature mortality related to PM<sub>2.5</sub> and O<sub>3</sub> in 2025

Mortality attributable to air pollution (deaths ×10 <sup>3</sup> )									
		PM <sub>2.5</sub>					O <sub>3</sub>	Total	
	Strata	Population (×10 <sup>6</sup> )	ALRI <5 yr	IHD ≥30 yr	CEV ≥30 yr	COPD ≥30 yr	LC ≥30 yr	COPD ≥30 yr	
Africa	Afr-D	538	102	60	99	14	2	5	282
	Afr-E	597	17	30	27	4	1	2	81
Americas	Amr-A	395	0	48	8	5	7	6	74
	Amr-B	561	0	9	3	1	1	3	17
	Amr-D	105	0	1	0	0	0	0	1
Eastern Mediterranean	Emr-B	197	2	55	33	3	2	5	100
	Emr-D	579	61	133	109	17	6	19	345
Europe	Eur-A	423	0	79	37	8	17	5	146
	Eur-B	246	1	74	46	4	9	4	138
	Eur-C	221	0	123	34	3	5	0	165
Southeast Asia	Sear-B	362	3	54	38	11	5	7	118
	Sear-D	1,697	76	461	398	201	21	155	1,312
Western Pacific	Wpr-A	158	0	11	11	1	5	4	32
	Wpr-B	1,760	18	377	1,038	284	138	51	1,906
World		7,838	280	1,515	1,881	556	219	266	4,717



Extended Data Table 5 | Premature mortality related to PM<sub>2.5</sub> and O<sub>3</sub> in 2050

Mortality attributable to air pollution (deaths ×10 <sup>3</sup> )									
		PM <sub>2.5</sub>					O <sub>3</sub>	Total	
	Strata	Population (×10 <sup>6</sup> )	ALRI <5 yr	IHD ≥30 yr	CEV ≥30 yr	COPD ≥30 yr	LC ≥30 yr	COPD ≥30 yr	
Africa	Afr-D	874	137	121	203	28	4	8	501
	Afr-E	933	21	64	58	10	1	4	158
Americas	Amr-A	451	0	59	10	6	10	7	92
	Amr-B	609	0	15	5	1	1	4	26
	Amr-D	131	0	1	0	0	0	0	1
Eastern Mediterranean	Emr-B	222	2	80	50	5	3	7	147
	Emr-D	799	64	241	196	32	10	33	576
Europe	Eur-A	431	0	85	43	9	20	5	162
	Eur-B	252	1	95	68	6	12	5	187
	Eur-C	203	0	127	45	3	5	1	181
Southeast Asia	Sear-B	382	3	69	51	14	7	8	152
	Sear-D	1,950	101	796	755	405	41	219	2,317
Western Pacific	Wpr-A	149	0	10	9	0	4	4	27
	Wpr-B	1,712	16	403	1,110	309	151	53	2,042
World		9,098	346	2,166	2,604	828	270	358	6,572

**Extended Data Table 6 | Population and premature mortality (deaths per year) related to PM<sub>2.5</sub> and O<sub>3</sub> in the most polluted megacities and conurbations in 2010, 2025 and 2050**

	2010		2025		2050	
Megacity	Population (×10 <sup>6</sup> )	Deaths (×10 <sup>3</sup> )	Population (×10 <sup>6</sup> )	Deaths (×10 <sup>3</sup> )	Population (×10 <sup>6</sup> )	Deaths (×10 <sup>3</sup> )
London	8.1	2.8	9.1	3.4	10.2	4.2
Paris	8.4	3.1	9.2	3.8	10.2	4.6
Moscow	14.9	8.6	14.4	10.8	13.1	11.7
Po Valley	3.4	1.3	3.4	1.4	3.2	1.4
Istanbul	11.1	5.6	13.2	8.5	14.5	13.2
Teheran	9.7	2.9	11.1	4.8	11.4	6.9
Cairo	12.5	6.0	15.9	8.2	19.8	11.4
Lagos	8.3	3.7	12.7	6.0	22.0	11.2
Johannesburg <sup>1</sup>	6.9	1.5	7.7	2.3	8.6	3.8
Karachi	11.9	7.3	15.4	11.4	19.4	17.9
Mumbai <sup>2</sup>	18.0	10.2	22.1	17.4	26.8	33.1
Delhi	22.5	19.7	27.8	31.1	33.3	52.0
Kolkata	20.3	13.5	28.4	26.6	38.8	54.8
Dhaka	22.8	13.1	31.2	26.4	38.2	49.9
Jakarta	22.5	10.4	26.1	16.4	29.0	22.1
Chengdu	6.2	7.4	6.4	9.5	5.9	9.7
Beijing	10.8	13.7	11.3	17.3	10.4	17.7
Tianjin	3.7	4.9	3.9	6.2	3.6	6.3
Shanghai	14.1	14.9	14.3	18.9	13.2	19.4
Seoul	20.8	6.6	21.7	8.5	20.3	8.7
Tokyo	29.2	6.0	28.1	6.4	24.2	5.4
Osaka	13.5	2.8	12.8	3.1	10.9	2.6
Hong Kong	6.9	2.6	7.6	3.7	8.8	4.4
Pearl River area	53.1	49.2	56.0	65.2	52.9	67.4
Manila	19.8	0.6	26.5	2.3	37.3	4.5
Bangkok	8.8	3.1	9.5	4.9	9.2	5.7
New York	12.5	3.2	14.5	4.2	17.5	5.2
Los Angeles	12.2	4.1	14.6	5.2	17.7	7.0
Mexico City	10.7	1.6	12.3	3.3	13.9	5.3

<sup>1</sup> Includes Pretoria<sup>2</sup> Includes east suburb

The names of the megacities cities have been colour coded according to the WHO regions: Europe, black font; Eastern Mediterranean, blue; Africa, red; Southeast Asia, green; Western Pacific, brown; Americas, purple.

Extended Data Table 7 | Premature mortality related to PM<sub>2.5</sub> and O<sub>3</sub> for the population aged <5 years and ≥30 years

WHO region	Year	Population (×10 <sup>6</sup> )	Mortality attributable to air pollution (deaths ×10 <sup>3</sup> )						Total
			PM <sub>2.5</sub>					O <sub>3</sub>	
			ALRI <5 yr	IHD ≥30 yr	CEV ≥30 yr	COPD ≥30 yr	LC ≥30 yr	COPD ≥30 yr	
Africa	2010	809	90	55	77	11	2	2	237
	2010*			37					219
Americas	2010	930	0	44	8	4	7	5	68
	2010*			35					59
Eastern Mediterranean	2010	602	56	115	86	12	5	12	286
	2010*			104					275
Europe	2010	867	1	239	95	13	27	6	381
	2010*			213					355
Southeast Asia	2010	1,762	64	327	250	124	15	82	862
	2010*			169					704
Western Pacific	2010	1,812	19	299	794	209	107	35	1,463
	2010*			250					1,414
World	2010	6,783	230	1,079	1,311	374	161	142	3,297
	2010*			808					3,026

\*In these rows, IHD mortality related to residential energy use (RCO) and biomass burning has been excluded.

# Non-adaptive plasticity potentiates rapid adaptive evolution of gene expression in nature

Cameron K. Ghalambor<sup>1,2</sup>, Kim L. Hoke<sup>1,2</sup>, Emily W. Ruell<sup>1</sup>, Eva K. Fischer<sup>1</sup>, David N. Reznick<sup>3</sup> & Kimberly A. Hughes<sup>4</sup>

Phenotypic plasticity is the capacity for an individual genotype to produce different phenotypes in response to environmental variation<sup>1</sup>. Most traits are plastic, but the degree to which plasticity is adaptive or non-adaptive depends on whether environmentally induced phenotypes are closer or further away from the local optimum<sup>2–4</sup>. Existing theories make conflicting predictions about whether plasticity constrains or facilitates adaptive evolution<sup>4–12</sup>. Debate persists because few empirical studies have tested the relationship between initial plasticity and subsequent adaptive evolution in natural populations. Here we show that the direction of plasticity in gene expression is generally opposite to the direction of adaptive evolution. We experimentally transplanted Trinidadian guppies (*Poecilia reticulata*) adapted to living with cichlid predators to cichlid-free streams, and tested for evolutionary divergence in brain gene expression patterns after three to four generations. We find 135 transcripts that evolved parallel changes in expression within the replicated introduction populations. These changes are in the same direction exhibited in a native cichlid-free population, suggesting rapid adaptive evolution. We find 89% of these transcripts exhibited non-adaptive plastic changes in expression when the source population was reared in the absence of predators, as they are in the opposite direction to the evolved changes. By contrast, the remaining transcripts exhibiting adaptive plasticity show reduced population divergence. Furthermore, the most plastic transcripts in the source population evolved reduced plasticity in the introduction populations, suggesting strong selection against non-adaptive plasticity. These results support models predicting that adaptive plasticity constrains evolution<sup>6–8</sup>, whereas non-adaptive plasticity potentiates evolution by increasing the strength of directional selection<sup>11,12</sup>. The role of non-adaptive plasticity in evolution has received relatively little attention; however, our results suggest that it may be an important mechanism that predicts evolutionary responses to new environments.

A long-standing problem in evolutionary biology is to understand the relationship between environmentally induced variation observed within a generation, and genetically-based evolutionary changes between generations<sup>1–5</sup>. It has long been recognized that the expression of traits is plastic—the same genotype can produce a range of phenotypes in response to different environmental cues. However, the causal relationship between a trait's plasticity and that trait's evolution remains an unresolved and contentious problem<sup>1</sup>. Traditional models of adaptive evolution ignored any role for plasticity, because environmentally induced plasticity was viewed as non-heritable variation<sup>1–3</sup>. Current models recognize that environments can cause predictable patterns of plasticity that are either adaptive or non-adaptive with respect to the local phenotypic optimum; such plasticity may influence evolutionary change by altering the distribution of phenotypes upon which selection acts. For example, plasticity is adaptive when the phenotype is altered in the same direction favoured by natural selection in that environment<sup>4–12</sup>. Some models predict that adaptive plasticity weakens the strength of directional selection and slows adaptive

evolution<sup>6–8,13</sup>. Other models suggest that adaptive plasticity is a critical first step in the process of adaptive evolution (for example, via genetic assimilation or accommodation)<sup>1,14</sup>, for instance by increasing population persistence in new environments (the Baldwin effect) and allowing more time for selection to act on heritable variation<sup>4–10</sup>. In contrast, plasticity is non-adaptive when a population encounters an environment that induces the production of phenotypes further away from the local optimum<sup>4–12</sup>, resulting in a negative relationship between the direction of plasticity and the direction of adaptive evolution. Non-adaptive plasticity reduces relative fitness and is predicted to increase the strength of directional selection because traits are further from the phenotypic optimum, resulting in an evolutionary response sometimes referred to as 'genetic compensation' or 'counter-gradient variation'<sup>11,12</sup>. Laboratory selection experiments have found support for a positive (adaptive)<sup>14,15</sup> and negative (non-adaptive)<sup>16</sup> relationship between the direction of plastic responses and the direction of evolution. However, testing such relationships in natural populations has been challenging because comparisons between ancestral and derived populations typically occur long after the populations have diverged<sup>17–20</sup>. Here, we test the relationship between plasticity and the early stages of evolutionary divergence using experiments in nature. We assess both ancestral plasticity in the source population and evolved changes in replicated derived populations by comparing plastic and evolved patterns of gene expression.

We quantified gene expression in Trinidadian guppies derived from natural populations and from populations undergoing early divergence following an experimental translocation. Individuals from a population that experiences high mortality from fish predators (high-predation, denoted as HP), particularly the pike cichlid (*Crenicichla frenata*), were introduced into each of two low-predation sites lacking cichlids: 'Intro1' and 'Intro2' (Extended Data Fig. 1). Thirty-eight gravid females and 38 mature males were introduced into each stream. One year after the introduction (3–4 guppy generations), guppies were collected from the ancestral HP source population, descendant introduction populations (Intro1 and Intro2), and a naturally colonized low-predation guppy population (denoted as LP) from the same drainage (Methods). The natural LP population represents an older evolutionary descendant of the HP source population<sup>21</sup> adapted to the same predation regime as the experimental populations. It thus provides an a priori prediction for the expected direction of evolutionary change.

To assess plastic and evolved changes in transcription, we bred wild-caught fish under common laboratory conditions for two generations and generated unique family lines within each of the four populations. Two generations of rearing in a common environment controls for environmental, maternal and other non-heritable sources of variation. Within 24 h of birth, second generation full-siblings of each family were randomly split between tanks that differed in exposure to chemical predator cues. Siblings reared with predator cues were raised in recirculating units that housed a cichlid within the water supply<sup>22</sup>. Cichlids were fed two guppies per day. Predator cues included both

<sup>1</sup>Department of Biology, Colorado State University, Fort Collins, Colorado 80523, USA. <sup>2</sup>Graduate Degree Program in Ecology, Colorado State University, Fort Collins, Colorado 80523, USA. <sup>3</sup>Department of Biology, University of California, Riverside, California 92521, USA. <sup>4</sup>Department of Biological Science, Florida State University, Tallahassee, Florida 32306-4295, USA.



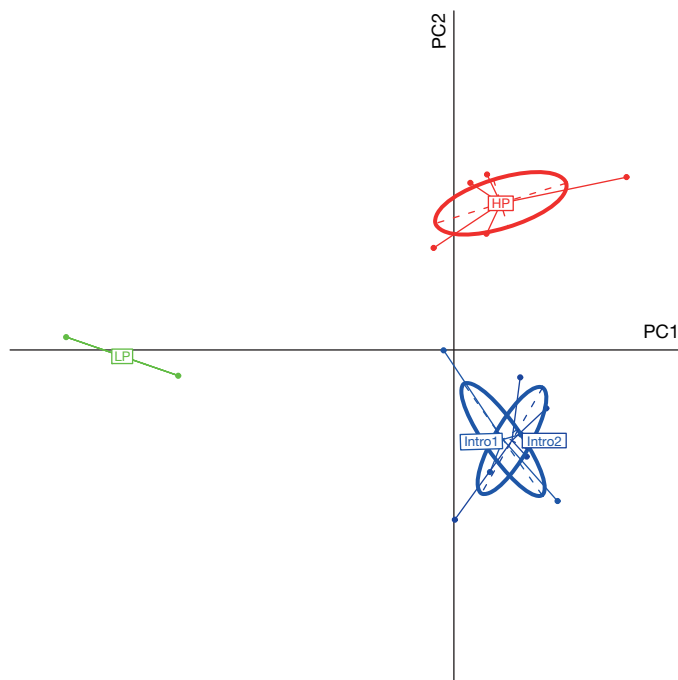
predator kairomones from the cichlid as well as any alarm pheromones from guppies, simulating the ancestral olfactory environment<sup>22</sup>. Guppies reared without predator cues were housed in identical recirculating units without the cichlid predator, simulating the derived environment. Differences in transcription between siblings reared in these two environments represent predator-induced plasticity in gene expression, while differences between populations measured under the same conditions for multiple generations represent heritable differences<sup>23</sup>.

To determine whether the introduction populations showed evidence for adaptive evolutionary divergence, we measured patterns of transcription in all four populations under the derived rearing environment. We measured the abundance of 37,493 messenger RNA transcripts expressed in whole brains of mature males reared without predator cues (mean age = 124.03 days old, range = 118–154), using high-throughput RNA sequencing. We used multivariate between-group principal components analysis (Methods) to visualize overall transcription differences among the four populations (Fig. 1). Two major axes explained 74.5% of the variation. Principal component 1 (PC1; 44.4% of variation) separated the naturally occurring LP population from the natural HP and introduction populations, and thus appears to reflect long-term divergence between these populations. PC2 (30.1% of variation) separated the HP source population from the two introduction populations and the natural LP population, thus capturing a signal of rapid and parallel evolutionary divergence to the LP environment (Fig. 1). Whereas genetic drift, founder effects, and unique attributes of each of the introduction streams would be expected to produce independent genetic changes in the introduction populations<sup>24</sup>, the parallel change of Intro1 and Intro2 in the same direction as the natural LP population supports the interpretation that PC2 describes rapid adaptive evolution. Indeed, the rate of evolutionary divergence in gene expression between the source population and

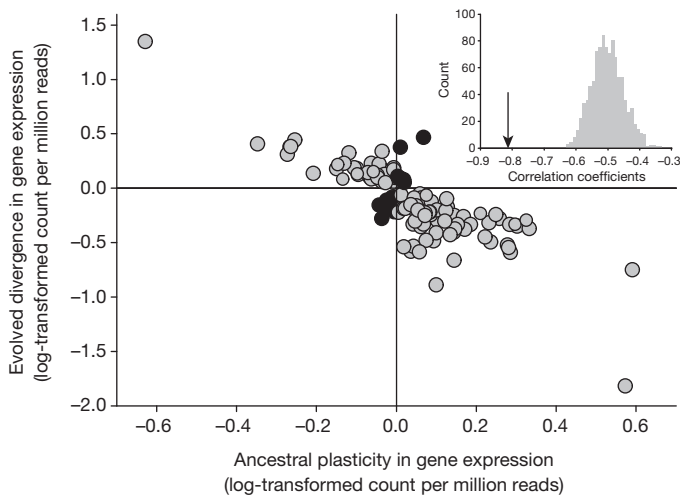
introduction populations for the top 500 transcripts loading on PC2 (median Haldanes (a change in phenotypic standard deviations per generation) in Intro1 = 0.256 and Intro2 = 0.226) are comparable to rapid rates of evolution observed in life history and morphology during previous experimental introductions of guppies<sup>25,26</sup> (Methods and Extended Data Fig. 2a, b).

To distinguish transcripts that exhibited evolution in the introduction populations as a result of selection from those that exhibited changes as a result of other processes, we identified transcripts that exhibited highly significant parallel evolutionary change in both introduction populations and that diverged in the same direction in the natural LP population. Permuted data sets ( $n = 250$ ) were generated by randomly reassigning population labels to individual samples. We then used general linear statistical models to assess divergence in the two introduction populations and the natural LP population (that is, HP versus Intro1 and HP versus Intro2 and HP versus LP) for each transcript (Methods). If the test statistic for each of the three contrasts fell in the extreme 5% of the distribution of the permutation test statistics, and the contrasts all had the same sign, we called the transcript concordantly differentially expressed (CDE). We found 135 transcripts that met these stringent criteria, which was many more than observed in the permuted data sets (median = 6, interquartile range = 3–14; Methods and Supplementary Table 1). By contrast, only one transcript diverged significantly in opposite directions in the two descendant introduction populations, consistent with expectations based on the distribution of permuted values (median = 1, interquartile range = 1–2). These 135 CDE transcripts loaded highly on PC2 (the median rank of the PC2 loadings for the CDE transcripts was 361 out of 37,493 total transcripts). The prevalence of these parallel changes suggests that this subset of transcripts evolved through the direct or indirect effects of natural selection, because genetic drift would have produced discordant as well as concordant evolution in the descendant introduction populations. Indeed, divergence in transcription between the ancestral and introduction populations greatly exceeded allele frequency divergence in putatively neutral microsatellite loci<sup>24</sup> (Extended Data Table 1). Collectively, these results demonstrate rapid and repeatable patterns of adaptive evolutionary divergence in transcription, similar to what has been observed for other fitness-related guppy traits following the colonization of low-predation environments<sup>22,26–28</sup>.

Given the evidence for rapid evolution of transcription, we determined if the pattern of ancestral plasticity in the HP source population predicted adaptive evolution in the descendant introduction populations. We assessed plasticity in the HP population by measuring the change in transcript abundance of full siblings reared with and without the predator cue (that is, simulating the ancestral high-predation and derived low-predation environments). If plasticity in transcript abundance was in the same direction as the parallel divergence observed in CDE transcripts, we considered plasticity to be adaptive. If the plastic changes were in the opposite direction as the evolved changes in CDE transcripts, we considered the plasticity to be non-adaptive (see Extended Data Fig. 3). We found a robust pattern of non-adaptive plasticity predicting evolutionary change in CDE transcripts; when HP fish were reared without the predator cue, the change in transcript abundance was overwhelmingly in the opposite direction to that of evolved changes in the descendant introduction populations (Fig. 2). The negative association between the direction of plasticity and the direction of evolution was highly significant ( $\chi^2 = 89.9$ , d.f. = 1), which is outside the range of all 250 permuted  $\chi^2$  values (range = 0.0–55.9), with 89% (120 of 135) of all transcripts exhibiting a plastic response opposite to the direction of evolution (see grey points in Fig. 2). Of the remaining 11% (15 of 135) of transcripts, when the direction of plasticity and evolution aligned, the degree of plasticity was negligible (see black points in Fig. 2). The correlation between ancestral plasticity and evolution ( $r = -0.82$ ) is substantially more negative than correlations generated from a randomization test



**Figure 1 | Rapid evolutionary divergence in gene expression as measured in second-generation laboratory-born guppies derived from the wild.** Shown is a principal components analysis of all 37,493 expressed genes in the four populations. HP is a naturally occurring high-predation population that is the source population for the two experimentally introduced populations, Intro1 and Intro2. LP is a naturally occurring low-predation population. Points represent individual families within each population, and are connected by solid lines. Dashed lines represent the major and minor axes of the confidence ellipse for each population.

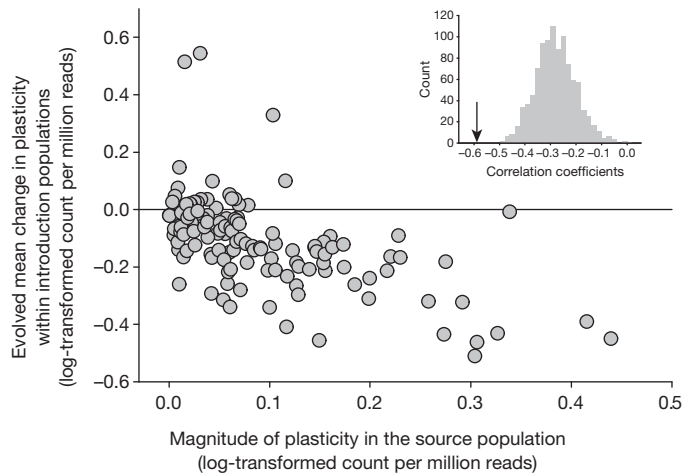


**Figure 2 | Rapid evolutionary divergence is highly correlated with non-adaptive plasticity.** Shown is a scatter plot of ancestral plasticity (change in transcript abundance to the absence of cichlid predator cues) against adaptive evolutionary divergence (135 concordantly differentially expressed transcripts) in the descendant populations transplanted to streams lacking cichlid predators. Grey points denote transcripts exhibiting non-adaptive plasticity, and black points denote adaptive plasticity. Inset shows the distribution of the Spearman rank correlations between evolutionary divergence and ancestral plasticity from 1,000 permuted correlation values for the 135 concordantly differentially expressed transcripts, with the arrow indicating the observed correlation, which is substantially more negative than all permuted values.

$P < 0.0001$ ; Fig. 2). These results suggest that plasticity potentiates rapid adaptive evolution, but not because plasticity is adaptive, as is assumed in many evolutionary models, but rather because it is non-adaptive and under stronger selection to change (Fig. 2). The same pattern is observed when we restrict the analysis to a separate data set that included 565 transcripts exhibiting significant plasticity to the rearing treatments in the HP source population (Supplementary Table 2 and Extended Data Fig. 4)

The magnitude of plasticity can also evolve in response to selection<sup>1–10</sup>. If natural selection acts most strongly on transcripts exhibiting non-adaptive plasticity, we predicted plasticity should evolve to be reduced in the descendant introduction populations. We tested this prediction by comparing plasticity in the ancestral source population to that in the derived introduction populations for the subset of transcripts that were CDE. The magnitude of plasticity decreased in the introduction populations (median change =  $-11\%$ , sign test  $M = -45.5$ ,  $P < 0.001$ ). Moreover, the decline in plasticity in these descendant populations was negatively associated with the magnitude of ancestral plasticity ( $P < 0.001$  based on a randomization test; Fig. 3), in accord with the idea that selection acts more strongly to decrease plasticity in those transcripts showing the greatest non-adaptive plasticity. Thus, traits exhibiting initially non-adaptive plastic responses to new environments may be a transient phenomenon, because selection may act to rapidly reduce their magnitude.

Attempts to model the effects of plasticity on subsequent adaptive evolution often assume that plasticity is adaptive. However, when populations experience novel environments, as when we experimentally transplanted guppies, many of the initial plastic responses are likely to be non-adaptive, because selection has not had an opportunity to act on the genetic variation for plasticity<sup>2–4</sup>. In such cases, both adaptive and non-adaptive plastic responses would be expected by chance, but traits exhibiting adaptive plasticity should be under weaker directional selection relative to traits exhibiting non-adaptive plasticity and further from the new phenotypic optimum<sup>11,12</sup>. Indeed, both theoretical and empirical studies show that adaptive plasticity reduces directional selection<sup>7,13,27</sup>. While we were unable to directly estimate the strength of selection on transcript abundance phenotypes,



**Figure 3 | Rapid evolution of reduced plasticity.** Shown is a scatter plot of the absolute values for the magnitude of ancestral plasticity (the normalized difference in transcript abundance between the presence and absence of cichlid cues) against the change in plasticity between the source and introduction populations. Inset shows the distribution of the Spearman rank correlations between the magnitude of plasticity in the ancestral population and the change in plasticity in the introduction populations from 1,000 permuted correlation values for the 135 concordantly differentially expressed transcripts, with the arrow indicating the observed correlation, which is substantially more negative than all permuted values.

previous introduction experiments have demonstrated strong directional selection and rapid adaptation in response to low-predation environments<sup>25,26</sup>. If traits exhibiting non-adaptive plasticity are under stronger directional selection, then newly established populations will probably face a dual challenge if they are to persist and avoid extinction. First, they must overcome the fitness costs associated with strong directional selection on non-adaptive responses, including declines in population size; and second, they must harbour enough genetic variation to rapidly respond to selection<sup>9,10,28</sup>. Because heritable genetic variation for transcription appears to be common<sup>24</sup>, the potential for rapid adaptation may ameliorate one set of costs. However, other costs may be more difficult to avoid, as models suggest that population size, the distance a non-adaptive trait is from the local optimum, and the relationship of that trait to fitness will ultimately determine whether populations persist<sup>10,28</sup>. In the case of the introductions here, such costs may have been reduced, because individuals were transplanted to relatively more 'benign' conditions, such that high predator-induced mortality was replaced with increased competition, reduced food availability, and other environmental factors characterizing the low-predation streams<sup>29</sup>.

Understanding the role of phenotypic plasticity in adaptive evolution remains a contentious problem in evolutionary biology, in part because few studies have been able to capture the initial patterns of plasticity and subsequent adaptive divergence of traits in natural populations. Nevertheless, it is during the early stages of adaptive divergence that selection in new environments is likely to be strongest<sup>9,10,28</sup>, and when plasticity will either reduce or exacerbate the initial mismatch between the mean and optimal phenotypic responses<sup>6–10</sup>. Recent work in these same guppy populations documents a similar pattern in which non-adaptive plasticity potentiates a rapid evolution of growth rate<sup>22</sup>, suggesting a general pattern that extends to other phenotypic traits. While such results are consistent with many models of how selection acts on phenotypes<sup>6–10</sup>, the role of non-adaptive plasticity in adaptive evolution remains understudied, despite arguments that it may be a common, but cryptic, form of evolution<sup>11,12</sup>. More generally, understanding when and how plasticity affects evolutionary response is critical for predicting the short- and long-term effects of environmental change on organisms. Predictive evolutionary models of

phenotypic plasticity also have practical importance. For example, disease states within organisms respond plastically to treatments and also evolve, thus gene expression profiles can be used (as was done here) to predict how response to treatment influences disease progression<sup>30</sup>. Additional experimental evolution studies, especially those conducted in natural environments, will be critical for validating and parameterizing future models of how plasticity influences evolutionary change.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 4 April; accepted 3 August 2015.**

**Published online 2 September 2015.**

1. West-Eberhard, M. J. *Developmental Plasticity and Evolution* (Oxford Univ. Press, 2003).
2. Schmalhausen, I. I. *Factors of Evolution: the Theory of Stabilizing Selection* (Blakiston, 1949).
3. López-Maury, L., Marguerat, S. & Bahler, J. Tuning gene expression to changing environments: from rapid responses to evolutionary adaptation. *Nature Rev. Genet.* **9**, 583–593 (2008).
4. Ghalambor, C. K., McKay, J. K., Carroll, S. P. & Reznick, D. N. Adaptive versus non-adaptive phenotypic plasticity and the potential for contemporary adaptation in new environments. *Funct. Ecol.* **21**, 394–407 (2007).
5. Baldwin, J. M. *Development and Evolution* (Macmillan Company, 1902).
6. Ancel, L. W. Undermining the Baldwin expediting effect: does phenotypic plasticity accelerate evolution? *Theor. Popul. Biol.* **58**, 307–319 (2000).
7. Price, T. D., Qvarnström, A. & Irwin, D. E. The role of phenotypic plasticity in driving genetic evolution. *Proc. R. Soc. Lond. B* **270**, 1433–1440 (2003).
8. Paenke, I., Sendhoff, B. & Kawecki, T. J. Influence of plasticity and learning on evolution under directional selection. *Am. Nat.* **170**, E47–E58 (2007).
9. Lande, R. Adaptation to an extraordinary environment by evolution of phenotypic plasticity and genetic assimilation. *J. Evol. Biol.* **22**, 1435–1446 (2009).
10. Chevin, L.-M., Lande, R. & Mace, G. M. Adaptation, plasticity, and extinction in a changing environment: towards a predictive theory. *PLoS Biol.* **8**, e1000357 (2010).
11. Grether, G. F. Environmental change, phenotypic plasticity, and genetic compensation. *Am. Nat.* **166**, E115–E123 (2005).
12. Conover, D. O., Duffy, T. A. & Hice, L. A. The covariance between genetic and environmental influences across ecological gradients: reassessing the evolutionary significance of countergradient and cogradients. *Ann. NY Acad. Sci.* **1168**, 100–129 (2009).
13. Wright, S. Evolution in Medelian populations. *Genetics* **16**, 97–159 (1931).
14. Waddington, C. H. Genetic assimilation. *Adv. Genet.* **10**, 257–293 (1961).
15. Suzuki, Y. & Nijhout, H. F. Evolution of a polyphenism by genetic accommodation. *Science* **311**, 650–652 (2006).
16. Schaum, C. E. & Collins, S. Plasticity predicts evolution in marine algae. *Proc. R. Soc. Lond. B* **281**, 20141486 (2014).
17. Losos, J. B. *et al.* Evolutionary implications of phenotypic plasticity in the hindlimb of the lizard *Anolis sagrei*. *Evolution* **54**, 301–305 (2000).
18. Wund, M. A., Baker, J. A., Clancy, B., Golub, J. L. & Foster, S. A. A test of the “flexible stem” model of evolution: ancestral plasticity, genetic accommodation, and morphological divergence in the threespine stickleback radiation. *Am. Nat.* **172**, 449–462 (2008).
19. McCairns, R. J. & Bernatchez, L. Adaptive divergence between freshwater and marine sticklebacks: insights into the role of phenotypic plasticity from an integrated analysis of candidate gene expression. *Evolution* **64**, 1029–1047 (2010).
20. Scoville, A. G. & Pfrender, M. E. Phenotypic plasticity facilitates recurrent rapid adaptation to introduced predators. *Proc. Natl Acad. Sci. USA* **107**, 4260–4263 (2010).
21. Willing, E.-M. *et al.* Genome wide single nucleotide polymorphisms reveal population history and adaptive divergence in wild guppies. *Mol. Ecol.* **19**, 968–984 (2010).
22. Handelsman, C. A. *et al.* Predator-induced phenotypic plasticity in metabolism and rate of growth: rapid adaptation to a novel environment. *Integr. Comp. Biol.* **53**, 975–988 (2013).
23. Gibson, G. & Weir, B. The quantitative genetics of transcription. *Trends Genet.* **21**, 616–623 (2005).
24. Leder, E. H. *et al.* The evolution and adaptive potential of transcriptional variation in sticklebacks—Signatures of selection and widespread heritability. *Mol. Biol. Evol.* **32**, 674–689 (2015).
25. Reznick, D. A., Bryga, H. & Endler, J. A. Experimentally induced life-history evolution in a natural population. *Nature* **346**, 357–359 (1990).
26. Reznick, D. N., Shaw, F. H., Rodd, F. H. & Shaw, R. G. Evaluation of the rate of evolution in natural populations of guppies (*Poecilia reticulata*). *Science* **275**, 1934–1937 (1997).
27. Charmantier, A. *et al.* Adaptive phenotypic plasticity in response to climate change in a wild bird population. *Science* **320**, 800–803 (2008).
28. Gomulkiewicz, R. & Holt, R. D. When does evolution by natural selection prevent extinction? *Evolution* **49**, 201–207 (1995).
29. Reznick, D., Butler, M. J., IV & Rodd, H. Life-history evolution in guppies. VII. The comparative ecology of high- and low-predation environments. *Am. Nat.* **157**, 126–140 (2001).
30. Merlo, L. M., Pepper, J. W., Reid, B. J. & Maley, C. C. Cancer as an evolutionary and ecological process. *Nature Rev. Cancer* **6**, 924–935 (2006).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** This work was supported by grants from the National Science Foundation (DEB-0846175 to C.K.G., EF-0623632 to D.N.R., and IOS-0934451 and IOS-1354775 to K.A.H.). We thank C. Handelsman, K. Langin, D. Broder, E. Duval, I. Janowitz, E. Lange, A. Shah, J. Havrid, E. Kane and L. Angeloni for helpful comments on the study. Computing for this project was performed on the Spear cluster at the Research Computing Center at the Florida State University.

**Author Contributions** C.K.G., K.L.H. and K.A.H. planned and executed the study, E.W.R. reared the fish, E.K.F. collected the tissues K.A.H. analysed the gene expression data, D.N.R. planned and oversaw the field introduction experiments, and C.K.G. oversaw the laboratory experiments. All authors participated in writing the manuscript.

**Author Information** The sequence data are available at the Sequence Reads Archive (SRA) under accession number SRP06236. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to C.K.G. ([cameron1@colostate.edu](mailto:cameron1@colostate.edu)) or K.A.H. ([kahughes@bio.fsu.edu](mailto:kahughes@bio.fsu.edu)).



## METHODS

No statistical methods were used to predetermine sample size. The investigators were not blinded to allocation during experiments and outcome assessment.

**Study system and populations.** Guppies are a model system in evolutionary biology because they provide an opportunity to study rapid adaptive evolution in the wild<sup>25,26</sup>. In lowland rivers, guppies occur in diverse fish communities where they experience high mortality from a number of fish predators. In small upstream tributaries, guppies occur in simpler communities, typically co-existing only with the killifish (*Rivulus hartii*), which poses little risk to adult guppies resulting in a low predator-induced mortality rate<sup>25,26</sup>. Past research has shown that numerous life history, behavioural, and morphological traits vary between these contrasting environments, and that these differences can evolve rapidly following experimental introductions<sup>25,26</sup>. We sampled four populations of guppies within the Guanapo River drainage in the Northern Range Mountains of Trinidad, West Indies (Extended Data Fig. 1). The first population, hereafter referred to as HP, is a naturally occurring population subject to high predation in the lower Guanapo river drainage that contains a variety of predator species, including the common predator on guppies, the pike cichlid<sup>25,26</sup>. The second population, hereafter referred to as LP, represents a native low-predation population from the same drainage and was sampled from the upstream Taylor tributary of the Guanapo river, where guppies co-exist with only *R. hartii*. *R. hartii* are gape-limited omnivores that prey primarily on juvenile guppies<sup>25,26</sup>. The remaining two populations were experimentally established in two low-predation tributaries (the Lower Lalaja, and the Upper Lalaja) within the Guanapo drainage.

**Introduction experiments.** In March 2008, HP guppies were introduced into the Lower Lalaja (denoted as Intro1) and Upper Lalaja tributaries (denoted as Intro2) of the Guanapo drainage<sup>22</sup>. The two introduction populations were established in 100-m reaches of these small, first-order tributaries. The upper limit of the introduction reach on the Lower Lalaja was bounded by a waterfall, which was artificially enhanced to prevent emigration and the establishment of populations above the streams receiving introductions. The upper limit of the Upper Lalaja introduction reach had a natural barrier. The lower limit of both introduction sites had natural barriers, which blocked immigration from downstream populations of guppies. The streams below these downstream barriers were also guppy-free before our introduction and were separated from the main river by additional barriers.

Each stream was stocked with 38 gravid females and 38 mature males. These fish had been collected as juveniles, reared to maturity in single sex groups, and then mated in groups of 4–5 males and 4–5 females per breeding group before introduction. To minimize the potential for founder effects and equalize genetic diversity in each stream, males and females from each breeding group were introduced into alternate streams. Doing so increased the effective population size of each population, because females retained the sperm from mating with one set of 38 males, then were introduced and subsequently mated with a second set of 38 males. As part of a separate experiment the riparian forest canopy was experimentally thinned in the Intro2 stream before the introductions<sup>31</sup>, but the two introduction streams were similar in all other respects.

**Laboratory breeding experiments.** Laboratory populations used for the gene expression assays were second-generation laboratory fish that were originally derived from 30 adult females and 30 adult males collected from each of the HP, LP and two introduction populations (Intro 1, Intro 2) in March of 2009. This time period represented one year or 3–4 generations after the establishment of introduction populations. Fish were kept in 1.5-l tanks (Aquatic Habitats) connected to a custom-made recirculating system and maintained on a 12-h light cycle at  $25 \pm 1^\circ\text{C}$ <sup>22,32,33</sup>. Fish were reared on standardized food levels adjusted weekly for age and number of individuals per tank (morning, Tetramin tropical fish flakes, Spectrum Brands, Inc.; afternoon, brine shrimp (nauplii of *Artemia* spp.), Brine Shrimp Direct). The quantity of food offered daily approximated *ad libitum* and was comparable to the high level of food administered in other studies<sup>34</sup>.

We reared all wild-caught guppies for two generations under common garden conditions using a breeding design that retains the genetic variation of the original population, prevents inbreeding, and minimizes maternal and other environmental effects<sup>34</sup>. The first generation (G1) line in the laboratory was derived from wild-caught juveniles and reared to maturity in the lab. Wild-caught gravid females were housed individually until parturition and their offspring were used to create G1 family lines. Females that did not give birth within about 30–35 days of capture were randomly crossed with a wild-caught male; however, no two females were crossed with the same male. The G1 offspring from each brood were housed separately until sexed, and then separated into single-sex tanks. Juvenile females (28–56 days) can be identified by the presence of melanophores in a triangular patch that appears on their ventral abdomens, which is absent in males<sup>34</sup>. Sexing was accomplished by anaesthetizing guppies in buffered MS-222 (0.85 mg ml<sup>-1</sup>;

ethyl 3-aminobenzoate methane sulfonic acid salt) (Sigma-Aldrich) and observing the melanophores under a microscope. Males are considered to be sexually mature when the apical hood grows even with the tip of their gonopodium; females usually mature within  $\pm 1$ –2 days of males<sup>34</sup>. Mature males and females from each family line were then randomly chosen and crossed to other families to produce the second generation (G2). Each G2 family was the product of a unique cross, to minimize inbreeding and maximize the genetic variation within each population.

Within 24 h of birth, G2 full-sibling broods were randomly assigned to two 1.5-l tanks (2–10 full siblings per tank) that differed in exposure to chemical cues from a predator (reared with or without cues from a predator) using a split-brood design. Siblings reared with cues from predators were reared in recirculating units that housed a pike cichlid within the sump that supplied water to the tanks<sup>22,32,33</sup>. Chemical predation cues included both kairomones from the cichlid predator and alarm pheromones from the two guppies consumed daily by the cichlid. Guppies reared without cues from predators were housed in identical recirculating units without predators in the water supply. G2 juveniles were anaesthetized and sexed at 29 days (see above). From each population, we randomly selected 5–6 families to raise pairs of male siblings within each rearing treatment.

**RNA-sequencing.** Focal animals were euthanized by immersion in ice water followed by rapid decapitation (IACUC approved protocol #12-3818A). Whole brains were collected by cutting the head sagittally down the centre line and removing all brain tissue. Brains were then flash frozen in liquid nitrogen and stored at  $-80^\circ\text{C}$  until further processing. Tissue collection took  $<2$  min per fish, fast enough to minimize changes in gene expression due to handling. Whenever possible, we combined brains from two full-siblings in the same treatment group to ensure we could obtain sufficient RNA for sequencing, while minimizing variation among pooled individuals. To minimize temporal and circadian variation, we performed all dissections within 15 min after lights-on in the morning (fish were all kept on a 12:12 h light–dark cycle). In addition, gene expression levels at lights-on minimized expression differences in response to recent experiences. Our data thus represent baseline transcription levels. The age of the fish (118–154 days) and the timing of sampling were randomly distributed across populations. Because all dissections occurred within 15 min, no more than 8 individuals (1–2 families distributed in both treatments) could be sampled per day, and the order in which populations were sampled was randomized.

RNA was extracted from whole brain tissue using Qiagen RNeasy lipid extraction kit. A separate sequencing library was prepared for each pooled family, using unique index sequences from the Illumina Tru-Seq RNA kit following manufacturer's instructions. Sequencing libraries were constructed and sequenced on three lanes of an Illumina HiSeq 2000 at the HudsonAlpha Genomic Services Laboratory (Huntsville, Alabama) in April 2012. In total, 32 samples were sequenced in 5 lanes (sample sizes that passed quality filters:  $n = 5$  for HP reared with and without predators,  $n = 4$  for Intro1 and Intro2 reared with and without predators,  $n = 3$  for LP reared with predators, and  $n = 2$  reared without predators). We obtained 736,693,718 100-base pair (bp) reads that passed the machine quality filter, with 17,517,493 to 28,265,561-bp reads per sample, and average quality  $>35.6$  for all samples.

Sequencing reads were mapped to a high-quality brain-specific reference transcriptome for *P. reticulata*. We constructed the reference from a data set containing  $>450$  million 100-bp paired-end reads, which were filtered for high-quality sequences and normalized *in silico* to compress the range in *k*-mer abundance. We used SeqMan NGEN 4.1.2 to perform the assembly, which contained 41,347 contigs,  $N50 = 2,548$ , and recovered 63% of Tilapia (*Oreochromis niloticus*) Ensembl proteins (Release 70). Contigs from the assembly were annotated by blastx queries against SwissProt (database downloaded 6 October 2012), UniProt/Trembl (28 November 2012), and nr (11 December 2012). Default parameters were used in the blastx queries, with *e*-value cut-off of  $1 \times 10^{-4}$ .

Reads were mapped to the reference assembly using Bowtie 2 v2.0.0 on a server running Red Hat Enterprise Linux version 6.5. We used a seed size of 20 bp, with no mismatches allowed in the seed (run options: `-D 15 -R 2 -N 0 -L 20 -i S,1,0.75`). We retained mappings with quality scores  $>30$  ( $<0.001$  probability that the read maps elsewhere in the reference) and kept only contigs represented by  $\geq 1$  count per million reads in at least three samples. After removing low-abundance transcripts, 628,797,716 reads (85.3%) mapped to 37,493 unique contigs in the reference transcriptome. We used the number of reads mapping to each of those contigs along with TMM-normalized library sizes<sup>35</sup> to analyse differential expression.

**Data analysis.** Between-group analysis (BGA) was conducted<sup>36</sup> as implemented in the R package made4 (ref. 37). BGA is a multivariate discriminant approach that is appropriate when the number of variables exceeds the number of cases; it is carried out by ordinating groups of samples and projecting the individual sample locations on the resulting axes. We used principal components analysis (PCA) as the ordination method (Fig. 1). To quantify the rate of evolution along the axis separating the HP source population from the introduction populations, we cal-



culated evolutionary divergence in Haldanes<sup>38</sup>. We assumed a time of 3.5 generations, and used the difference in mean transcript abundance in the no predator treatment with a pooled standard deviation<sup>38</sup> (see Extended Data Figs 2a, b).

We used random permutation tests to evaluate differential expression across populations and treatment groups. Permuted data sets were generated by randomly reassigning entire RNA-seq samples among population and treatment categories to produce an empirical null distribution against which to test hypotheses. This approach preserves any non-independence among transcripts that could bias inferences if the non-independence were not taken into account. We first computed transcript-specific test statistics from the actual data (see below) and compared that statistic to the distribution of the same statistic derived from 250 permuted data sets. If the statistic for the real data fell within the extreme tails of the permuted values for that transcript, we called the transcript differentially expressed (DE). To determine if more transcripts were called DE than expected, we compared the number of DE transcripts in the real data set to the distribution of that number in the 250 permuted data sets.

To determine if transcripts were significantly evolved in each introduction population we restricted the analysis to samples collected from fish reared without predator cues. For both the actual and the permuted data sets, a general linear model was applied separately to each transcript, with the normalized transformed number of reads as the dependent variable and population (HP and Intro1 or HP and Intro2, depending on the analysis) as a fixed effect. We then used general linear statistical models to assess divergence in the two introduction populations and the natural LP population (that is, HP versus Intro1 and HP versus Intro2 and HP versus LP) for each transcript. If the test statistic for each of the three contrasts fell in the extreme 5% of the distribution of the permutation test statistics, and the contrasts all had the same sign, we called the transcript concordantly differentially expressed (CDE). To calculate the number of transcripts expected to be called CDE in the two introduction populations under random expectations, we conducted this same analysis in each of the 250 permuted data sets, and calculated the number of transcripts meeting the same criteria. This permutation analysis accounts for any spurious associations that might result from comparing both introduction populations to the same ancestral HP population<sup>39</sup>.

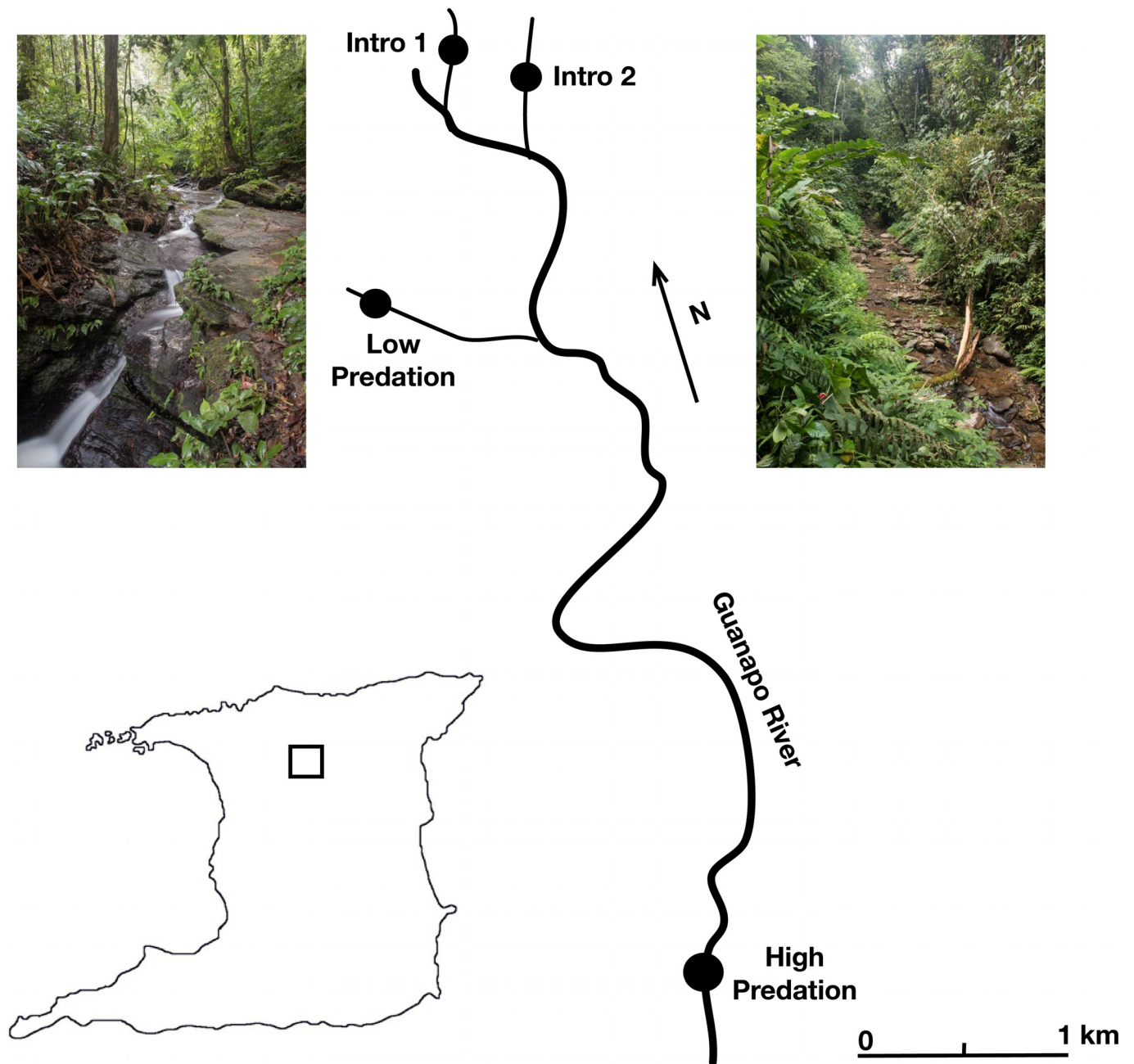
To test if the divergence in gene expression is greater than would be expected by neutral processes, we calculated  $P_{ST}$  (a measure of phenotypic divergence between populations) from phenotypic variance components as in ref. 40, assuming  $h^2 = 0.5$  (where  $h$  = heritability of a trait) for transcript expression level. This  $h^2$  estimate is substantially higher than the average estimate from a recent analysis in sticklebacks<sup>24</sup>, making our comparison of  $P_{ST}$  with published  $F_{ST}$  estimates ( $F_{ST}$  is a measure of genetic divergence between populations) conservative with respect to the hypothesis that divergence is greater than expected under genetic drift.

We assessed the association between evolutionary divergence and ancestral plasticity in gene expression by conducting likelihood ratio tests of independence and comparing the resulting  $\chi^2$  value to the distribution of  $\chi^2$  values produced by conducting the same test on the 250 permuted data sets. Similarly, for the CDE transcripts, we calculated the Spearman rank correlation between evolution (mean change in expression level between HP and introduction populations in the no-predator-cue environment) and plasticity (mean change in expression in the

HP ancestral population reared in the two predator-exposure environments), and compared that value to the distribution of values obtained from 1,000 random permutations of the population and treatment group labels. This permutation analysis accounts for any spurious correlation that can result because the calculations for evolutionary divergence and plasticity share a common term (mean expression level in the HP source population reared without predator cues)<sup>39</sup>.

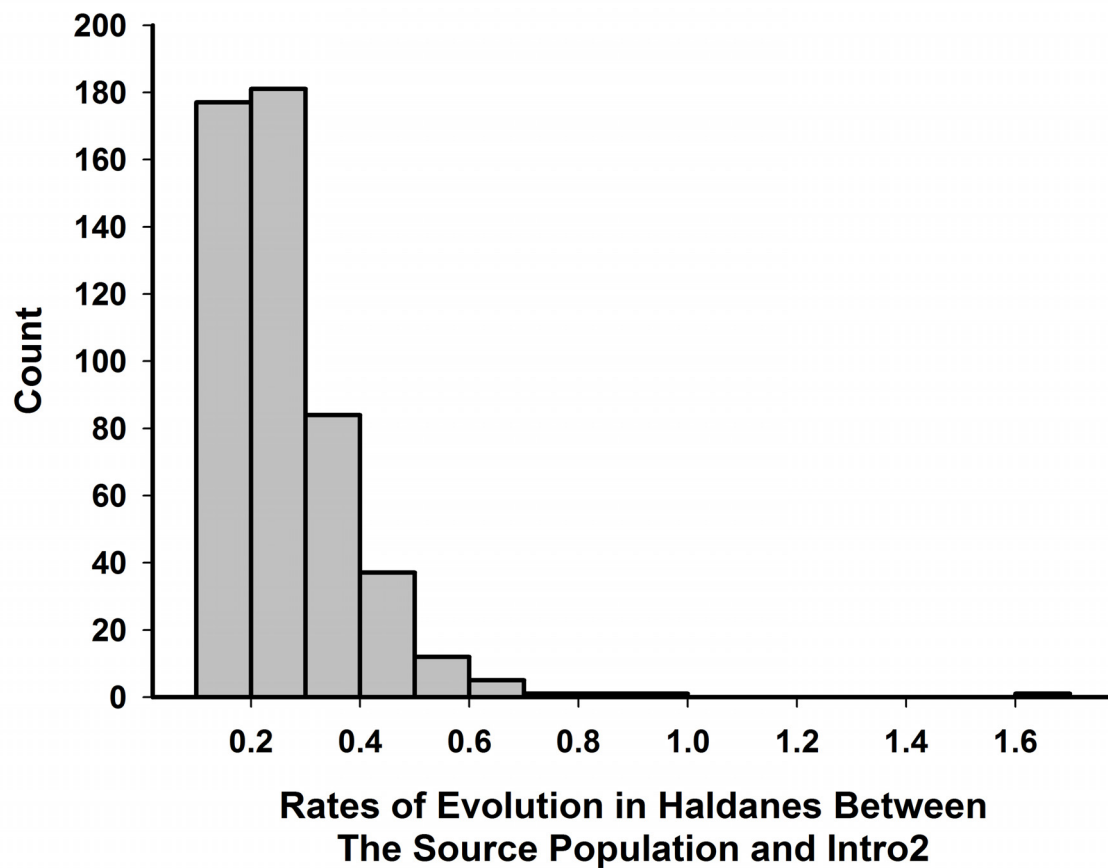
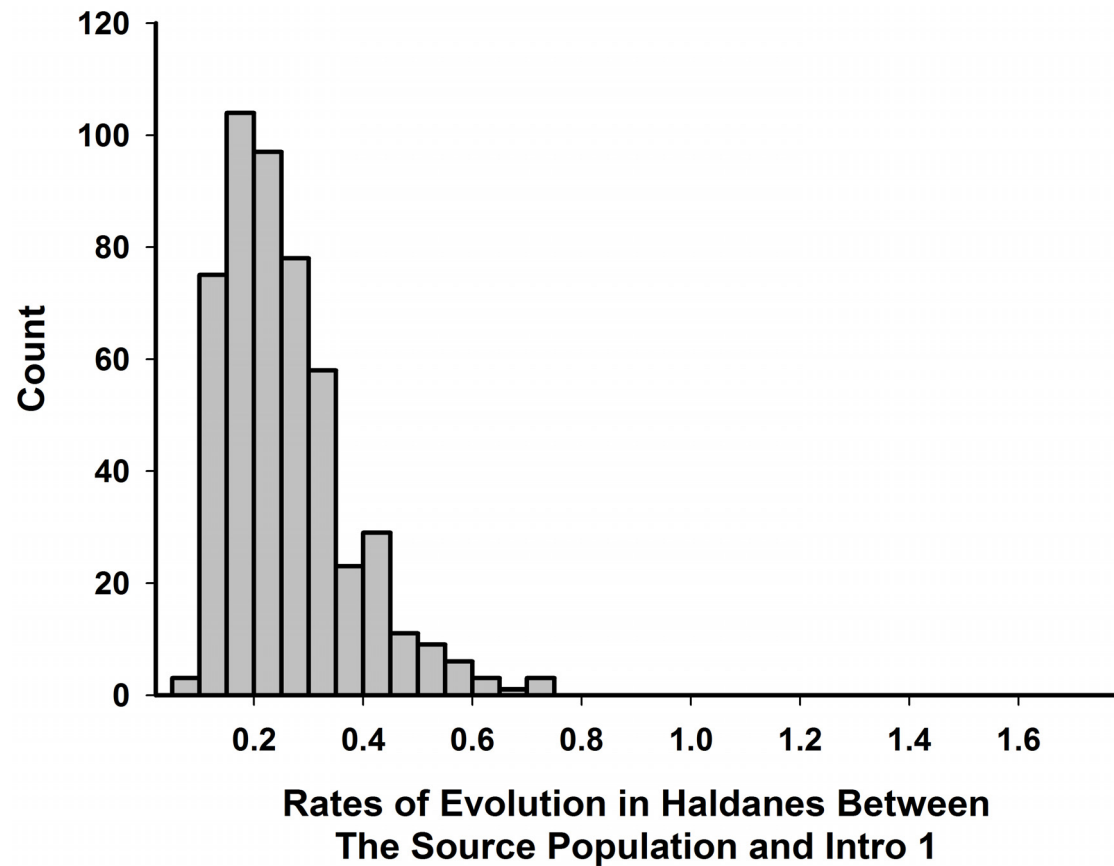
For CDE transcripts, we quantified plasticity in the source population and in the introduced populations as the difference in the mean expression values (normalized log-transformed number of reads mapping to a given transcript) for each transcript in the two predator-cue treatment groups within each population. We then calculated the change in these plasticity values between the source and introduction populations and used a nonparametric sign test to determine if that change was significant. We evaluated the association between ancestral and descendant plasticity in the CDE transcripts using a Spearman's rank correlation, and determined significance of that correlation using a random permutation test. Starting with the mean expression levels for each transcript within each population/treatment group, we randomly permuted the population/treatment labels 1,000 times, recalculated ancestral and derived plasticity values for each transcript in each permutation, and calculated Spearman's rank correlation of the permuted values. All statistical analyses were implemented in SAS 9.4 (SAS 2011) running in a Linux environment.

31. Kohler, T. J., Heatherly, T. N. II, El-Sabaawi, R. W., Zandonà, E., Marshall, M. C., Flecker, A. S., Pringle, C. M., Reznick, D. N. & Thomas, S. A. Flow, nutrients, and light availability influence Neotropical epilithon biomass and stoichiometry. *Freshwater Sci.* **31**, 1019–1034 (2012).
32. Torres-Dowdall, J., Handelsman, C. A., Reznick, D. N. & Ghalambor, C. K. Local adaptation and the evolution of phenotypic plasticity in Trinidadian guppies (*Poecilia reticulata*). *Evolution* **66**, 3432–3443 (2012).
33. Ruell, E. W. *et al.* Fear, food and sexual ornamentation: plasticity of colour development in Trinidadian guppies. *Proc. R. Soc. Lond. B* **280**, 20122019 (2013).
34. Reznick, D. The impact of predation on life history evolution in Trinidadian guppies: genetic basis of observed life history patterns. *Evolution* **36**, 1236–1250 (1982).
35. Robinson, M. D. & Oshlack, A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* **11**, R25 (2010).
36. Culhane, A. C., Perrière, G., Considine, E. C., Cotter, T. G. & Higgins, D. G. Between-group analysis of microarray data. *Bioinformatics* **18**, 1600–1608 (2002).
37. Culhane, A. C., Thioulouse, J., Perrière, G. & Higgins, D. G. MADE4: an R package for multivariate analysis of gene expression data. *Bioinformatics* **21**, 2789–2790 (2005).
38. Gingerich, P. D. Rates of evolution on the time scale of the evolutionary process. *Genetica* **112–113**, 127–144 (2001).
39. Jackson, D. A. & Somers, K. M. The spectre of “spurious” correlations. *Oecologia* **86**, 147–151 (1991).
40. Leinonen, T., Cano, J. M., Mäkinen, H. & Merilä, J. Contrasting patterns of body shape and neutral genetic divergence in marine and lake populations of threespine sticklebacks. *J. Evol. Biol.* **19**, 1803–1812 (2006).
41. Fitzpatrick, S. W., Gerberich, J. C., Kronenberger, J. A., Angeloni, L. M. & Funk, W. C. Locally adapted traits maintained in the face of high gene flow. *Ecol. Lett.* **18**, 37–47 (2015).

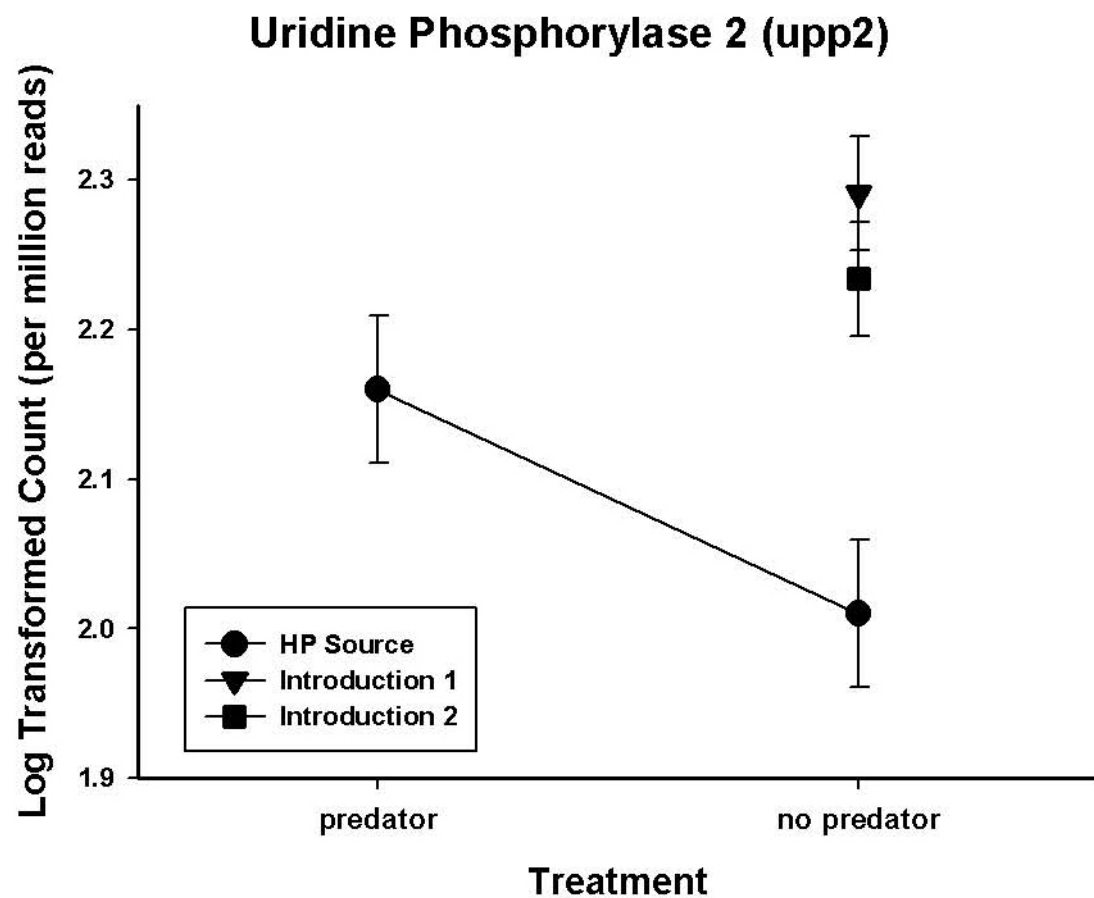


**Extended Data Figure 1 | Map of Trinidad where the experimental transplants took place.** Guppies were moved from a high-predation (HP) locality where they coexist with cichlid predators and introduced into two

streams that lacked cichlids and guppies, Intro1 (left photograph) and Intro2 (right photograph). A naturally occurring guppy population without cichlids, low-predation (LP), was sampled to provide a low-predation reference.



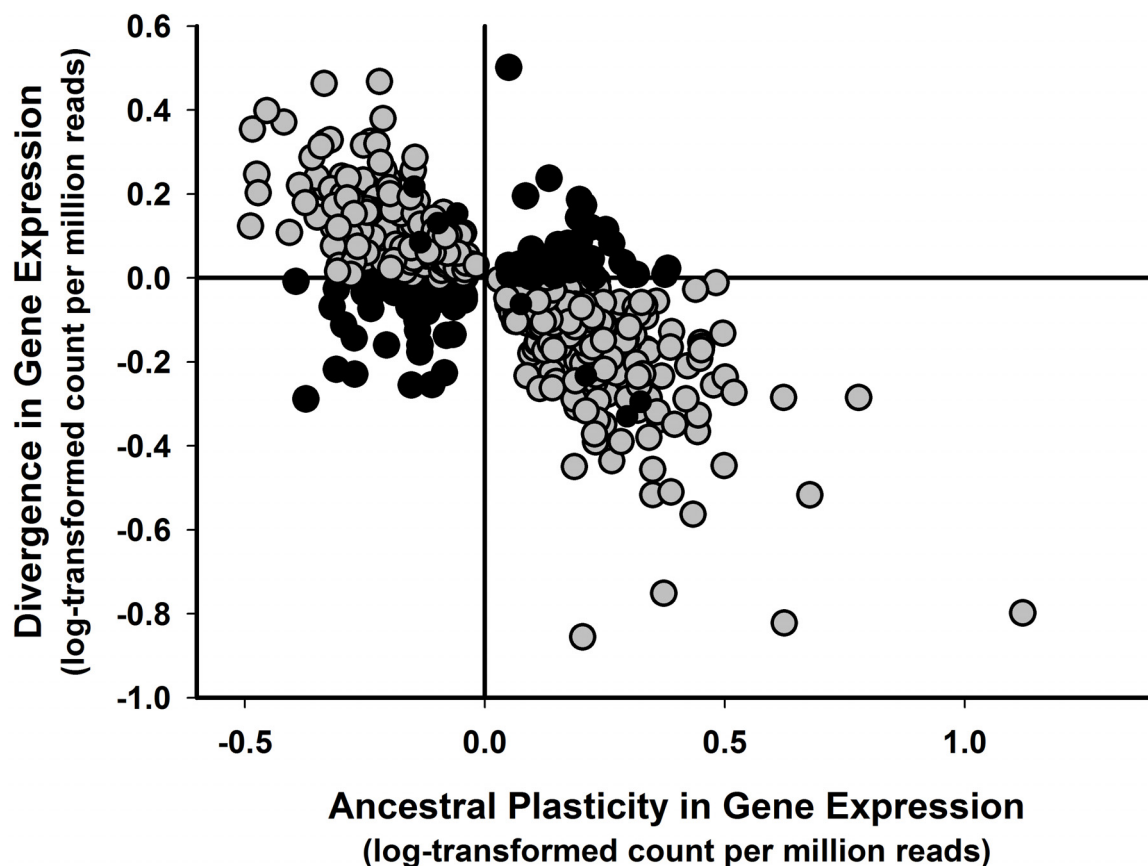
**Extended Data Figure 2 |** Frequency histogram of Haldanes for the top 500 transcripts loading on PC2—the axis representing rapid evolutionary divergence between the source and introduction populations. **a**, Intro1 (median Haldane = 0.256, range = 0.07–0.74). **b**, Intro2 (median = 0.226, range = 0.10–1.68).



**Extended Data Figure 3 | Ancestral plasticity and evolution in patterns of gene expression for a representative gene: uridine phosphorylase 2 (*upp2*).** Shown is the plastic response of the high-predation source population and the evolved responses in the two experimental introduction populations

(Intro1 and Intro2). In this case the plastic response results in a decrease in expression, whereas the evolved response in the introduction populations is to increase expression, thus illustrating non-adaptive plasticity.





**Extended Data Figure 4 | Scatter plot of ancestral plasticity (change in transcript abundance to the absence of cichlid predator cues) and population divergence.** Shown are the 565 transcripts that exhibited significant differences in expression between the predator and non-predator rearing treatments in the HP source population. We found a similar pattern as was found for the CDE transcripts (Fig. 2): 75% (424 out of 565) of the significantly plastic genes exhibited population divergence in the introduction populations

in the opposite direction of plasticity ( $\chi^2 = 284.2$ , d.f. = 1). This result falls in the upper percentile of the 250 permuted  $\chi^2$  values; median permuted values = 19.1, interquartile range = 6.7–50.8. Only eight transcripts were common to the data sets that were significantly evolved (CDE; Figs 2, 3) and significantly plastic, suggesting that short-term plastic responses and longer-term evolutionary responses involve largely different sets of genes.

**Extended Data Table 1 | Comparison of gene expression divergence ( $P_{ST}$ ) with divergence of putatively neutral microsatellite loci ( $F_{ST}$ )**

	Intro1 <sup>c</sup>	Intro2 <sup>d</sup>	
$P_{ST}$ <sup>a</sup>	0.32 (0.21)	0.27 (0.21)	Only CDE transcripts
	0.05 (0.11)	0.05 (0.12)	Only non-CDE transcripts
	0.05 (0.11)	0.05 (0.10)	All transcripts
$F_{ST}$ <sup>b</sup>	0.01	N/A	10 microsatellite loci

<sup>a</sup>Quantitative divergence estimated by  $P_{ST}$ , a phenotypic proxy for quantitative genetic divergence  $Q_{ST}$ <sup>40</sup>, calculated under the conservative assumption that half the within-population variation was heritable. Numbers in parentheses are standard deviations.  
<sup>b</sup>Neutral divergence estimated from 10 microsatellite loci<sup>41</sup>.  
<sup>c</sup>Divergence between the ancestral HP site (Guanapo) and the Intro1 site (Lower Lalaja).  
<sup>d</sup>Divergence between the ancestral HP site (Guanapo) and the Intro2 site (Upper Lalaja).

# A new cyanogenic metabolite in *Arabidopsis* required for inducible pathogen defence

Jakub Rajniak<sup>1</sup>, Brenden Barco<sup>2</sup>, Nicole K. Clay<sup>2</sup> & Elizabeth S. Sattely<sup>1</sup>

Thousands of putative biosynthetic genes in *Arabidopsis thaliana* have no known function, which suggests that there are numerous molecules contributing to plant fitness that have not yet been discovered<sup>1,2</sup>. Prime among these uncharacterized genes are cytochromes P450 upregulated in response to pathogens<sup>3,4</sup>. Here we start with a single pathogen-induced P450 (ref. 5), CYP82C2, and use a combination of untargeted metabolomics and coexpression analysis to uncover the complete biosynthetic pathway to 4-hydroxyindole-3-carbonyl nitrile (4-OH-ICN), a previously unknown *Arabidopsis* metabolite. This metabolite harbours cyanogenic functionality that is unprecedented in plants and exceedingly rare in nature<sup>6,7</sup>; furthermore, the aryl cyanohydrin intermediate in the 4-OH-ICN pathway reveals a latent capacity for cyanogenic glucoside biosynthesis<sup>8,9</sup> in *Arabidopsis*. By expressing 4-OH-ICN biosynthetic enzymes in *Saccharomyces cerevisiae* and *Nicotiana benthamiana*, we reconstitute the complete pathway *in vitro* and *in vivo* and validate the functions of its enzymes. *Arabidopsis* 4-OH-ICN pathway mutants show increased susceptibility to the bacterial pathogen *Pseudomonas syringae*, consistent with a role in inducible pathogen defence. *Arabidopsis* has been the pre-eminent model system<sup>10,11</sup> for studying the role of small molecules in plant innate immunity<sup>12</sup>; our results uncover a new branch of indole metabolism distinct from the canonical camalexin pathway, and support a role for this pathway in the *Arabidopsis* defence response<sup>13</sup>. These results establish a more complete framework for understanding how the model plant *Arabidopsis* uses small molecules in pathogen defence.

To identify cytochromes P450 potentially involved in the biosynthesis of novel defence-associated small molecules, we obtained raw data sets for all transcriptomics experiments dealing with biotic stress in *A. thaliana* from the NASCArrays database. We examined CYP genes present in the probeset and selected a candidate, CYP82C2, that is highly expressed under a variety of pathogen treatment conditions, but whose native function in *Arabidopsis* is unknown (Fig. 1a).

To identify small molecules whose levels change in a CYP82C2-dependent manner, we performed comparative metabolomics<sup>14</sup> with a homozygous transfer-DNA insertion line of CYP82C2. We used the bacterial pathogen *P. syringae* pv. *tomato* DC3000 harbouring the *avrRpm1* avirulence gene (*Psta*) as an elicitor since CYP82C2 expression is strongly upregulated 24 h after inoculation with this strain (Fig. 1a). We analysed tissue methanolic extracts of 11-day-old seedlings grown hydroponically in the presence of *Psta* by liquid chromatography–mass spectrometry (LC–MS), and computationally compared mutant and wild-type (WT) Col-0 metabolomes. From this analysis, we identified 11 compound mass signals that reproducibly and significantly differ between WT and *cyp82C2* (Fig. 1b); these mass ions are induced after pathogen elicitation and are not bacterially derived (Extended Data Fig. 1a).

We next sought to obtain clues about the structure of these compounds from their tandem mass spectra (MS/MS). MS/MS analysis revealed that the 11 compounds could be divided into two classes (A and B in Fig. 1b), assigned as indole-3-carboxaldehyde (IAL) deri-

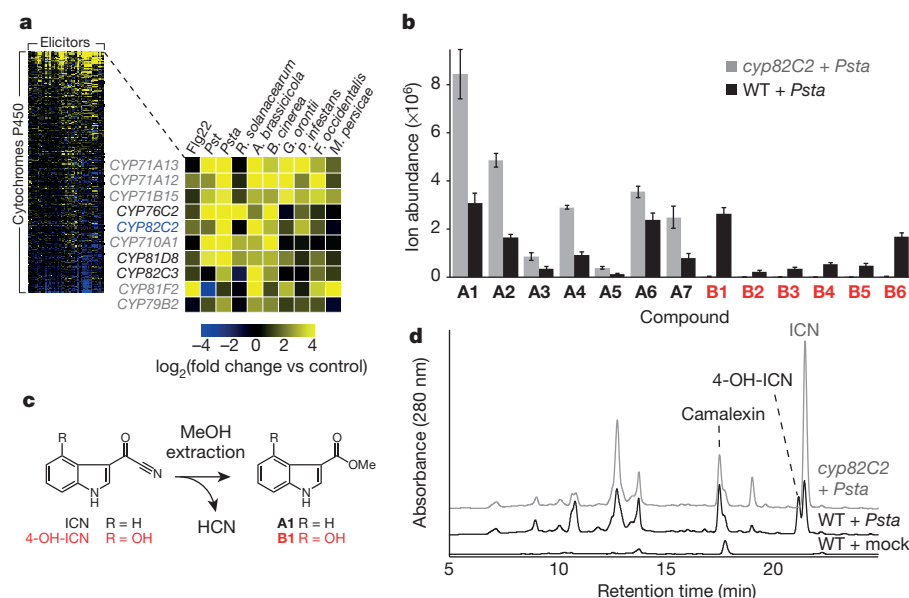
vatives with (B) and without (A) hydroxylated indole systems. Moreover, the fact that the *cyp82C2* mutant lacked all the hydroxylated derivatives but accumulated excess amounts of their non-hydroxylated counterparts suggested that CYP82C2 acts as an indolic hydroxylase. However, except for compound A1 (Fig. 2b), which was confirmed to be indole-3-carboxylic acid methyl ester, the structures of these compounds remained elusive.

To facilitate structural analysis, we investigated whether any of these compounds were exuded into the medium in the *cyp82C2* mutant seedling experiments (Fig. 1d). Filtered spent medium was loaded onto a C18 silica gel cartridge, and non-polar metabolites were eluted with acetonitrile and analysed by LC–MS. Surprisingly, the profile of spent medium extracted in this manner was notably different from that of tissue methanolic extracts: while small amounts of A2–A7 were present, no A1 could be detected; instead, a new ultraviolet-active compound with  $m/z = 171.0553$   $[M + H]^+$  dominated the LC–MS trace (Fig. 1d). NMR analysis of this compound followed by comparison with a synthetic standard established its identity as the novel metabolite indole-3-carbonyl nitrile (ICN) (Fig. 1c and Extended Data Fig. 2).

Chemically, the most striking feature of ICN is the presence of a highly reactive  $\alpha$ -ketonitrile moiety that, to our knowledge, has not been found in any plant natural product; however, benzoyl cyanide has been previously identified in the secretions of millipedes<sup>6,7</sup>. The  $\alpha$ -ketonitrile is susceptible to nucleophilic attack, resulting in the displacement of cyanide ion: in alkaline aqueous solution, ICN degrades to indole-3-carboxylic acid (ICA) (an alternative route to ICA in *Arabidopsis* has been reported<sup>15</sup>); in methanol, ICA methyl ester (A1) is formed instead, explaining the presence of A1 and the absence of ICN in methanolic extracts (Fig. 1c). Modifying the tissue extraction procedure by using an acidified 1:1 acetonitrile/water mixture enabled direct detection of ICN by LC–MS; additionally, when deuterated methanol was used, only the deuterated form of A1 was observed (Extended Data Fig. 1b–e). On the basis of its molecular formula and the synthesis of an authentic standard, A6 was shown to be a serine-ICN addition product (see Fig. 2b). However, in the presence of cysteine and structurally related compounds, ICN can undergo a spontaneous cycloaddition, resulting in the formation of a thiazoline ring and the net loss of ammonia. This last observation allowed us to determine the structures of and synthesize standards for compounds A2–A5, which are the cycloaddition products of ICN and cysteine (A4) or Cys–Gly dipeptide (A2) and their thiazole analogues (A5 and A3, respectively; see Fig. 2b, Extended Data Fig. 3, and Supplementary Table 1).

The absence of the hydroxylated analogues B1–B6 in the *cyp82C2* insertion line pointed to ICN as the likely substrate for this enzyme. Incubation of ICN with yeast-expressed CYP82C2 yielded only a trace amount of hydroxylated ICN, but a significant amount of 4-hydroxy-ICA (4-OH-ICA) (structure shown in Fig. 3), as confirmed by NMR spectroscopy and comparison with a synthetic standard (Extended Data Fig. 4a–d). Since CYP82C2 shows no activity on ICA, we deduced

<sup>1</sup>Department of Chemical Engineering, Stanford University, Stanford, California 94305, USA. <sup>2</sup>Department of Molecular, Cellular, and Developmental Biology, Yale University, New Haven, Connecticut 06511, USA.



**Figure 1 | Transcriptomic and metabolomic analyses implicate CYP82C2 in the biosynthesis of novel pathogen defence-related secondary metabolites.** **a**, Heat map of relative gene expression levels for cytochrome P450 genes in *Arabidopsis* under various pathogen stress conditions. The enlarged map shows the top 10 P450 genes after sorting by mean expression level over all conditions. Cytochromes P450 in grey have previously been biochemically characterized. **b**, Levels of the most significantly differing metabolites identified in seedling comparative metabolomics experiments with *cyp82C2*. Data represent the mean  $\pm$  s.d. of six biological replicates. **c**, ICA methyl ester (A1) and 4-OH-ICA methyl ester (B1) are methanolic degradation products of ICN and 4-OH-ICN. **d**, High-performance liquid chromatography traces of growth medium for WT and *cyp82C2* seedlings, showing *Psta*-dependent accumulation of ICN and 4-OH-ICN.

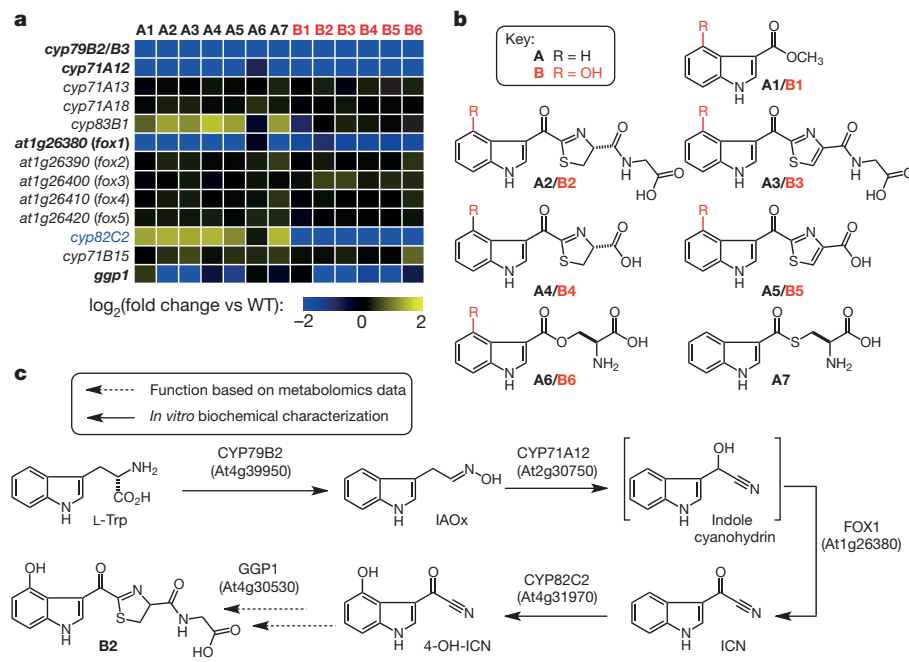
that CYP82C2 converts ICN to 4-OH-ICN, competing with hydrolysis of ICN to ICA (Extended Data Fig. 4e, f). Further experiments with chemically synthesized 4-OH-ICN showed that its half-life is approximately 3 min in aqueous solution at pH = 7.5 (Supplementary Table 2), rendering direct isolation of the 4-OH-ICN product infeasible. Chemical synthesis of 4-OH-ICN further enabled the synthesis of the 4-hydroxy derivatives of A1–A6, confirming that these correspond to compounds B1–B6 seen in WT tissue extracts (Extended Data Fig. 3 and Supplementary Table 1). Therefore, all the metabolites identified in our initial metabolomics experiment with *cyp82C2* are ultimately derived from ICN, whether as artefacts of the extraction (A1 and B1), or as *in vivo* addition products (A2–A7, B2–B6).

We next investigated the biosynthesis of ICN, using the CYP82C2 gene as bait for coexpression analysis. For our pathogen data set, the CYP79B2 gene, whose encoded enzyme converts tryptophan (Trp) into indole-3-acetaldoxime (IAOx)<sup>16</sup>, has the second highest correlation (Pearson's  $r$ ) with CYP82C2 among all genes profiled (Supplementary Table 3). We performed a metabolomic analysis of the *cyp79B2 cyp79B3*

double knockout line<sup>17</sup>, which is deficient in IAOx production. No ICN-derived metabolites are produced in this mutant (Fig. 2a), indicating that ICN is derived from IAOx.

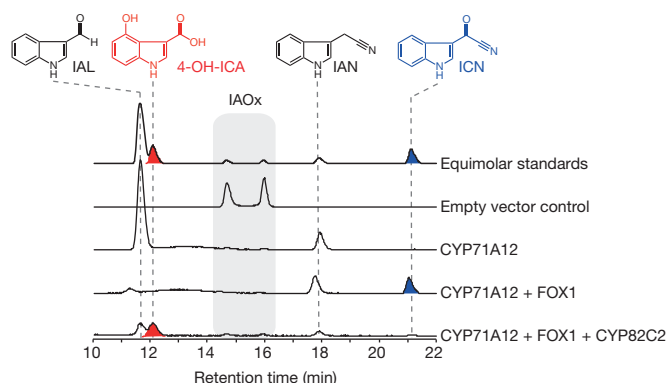
In searching for the enzyme(s) responsible for further conversion of IAOx to ICN, we postulated a biosynthetic route paralleling that of the cyanogenic glycoside dhurrin<sup>8</sup>: a CYP79-catalysed formation of an aldoxime, followed by a CYP71-catalysed formation of a cyanohydrin intermediate. In the dhurrin pathway, the cyanohydrin is glucosylated to yield the final product, whereas in ICN biosynthesis, a final dehydrogenation is required to produce an  $\alpha$ -ketonitrile (Fig. 2c).

Correlation analysis implicated CYP71A12, a P450 linked to camalexin biosynthesis<sup>18</sup>, as the most likely candidate gene for the cyanohydrin formation step (Supplementary Table 3). Profiling of the *cyp71A12* transfer-DNA insertion line, as well as transfer-DNA insertion lines of its two closest *Arabidopsis* homologues, CYP71A13 and CYP71A18, demonstrated that the CYP71A12 gene is in fact probably responsible: all ICN derivatives with the exception of A6 are at  $\sim$ 10% of WT levels in the *cyp71A12* mutant, but unaffected in the *cyp71A13*



**Figure 2 | Targeted metabolic profiling of candidate transfer-DNA insertion lines helps uncover the entire ICN biosynthetic pathway.** **a**, Heat map of mean ICN-derived metabolite levels relative to WT in *Psta*-elicited transfer-DNA insertion lines. Mutants in bold have significantly decreased levels of ICN derivatives. Note that A6 levels are not affected to the same extent as levels of other metabolites in any line except for *cyp79B2/B3*, hinting at an alternative biosynthetic route from IAOx for this metabolite. **b**, Structures of all ICN derivatives, confirmed by comparison with synthetic standards (see Extended Data Fig. 3 and Supplementary Table 1). **c**, Proposed biosynthetic pathway from Trp to 4-OH-ICN and downstream metabolites.





**Figure 3 | In vitro reconstitution of 4-OH-ICN biosynthesis from IAOx.** Combined extracted ion chromatograms (EICs) for IAOx substrate and reaction products for various subsets of enzymes in the 4-OH-ICN pathway; 4-OH-ICN could not be detected directly and its hydrolysis product 4-OH-ICA is shown instead.

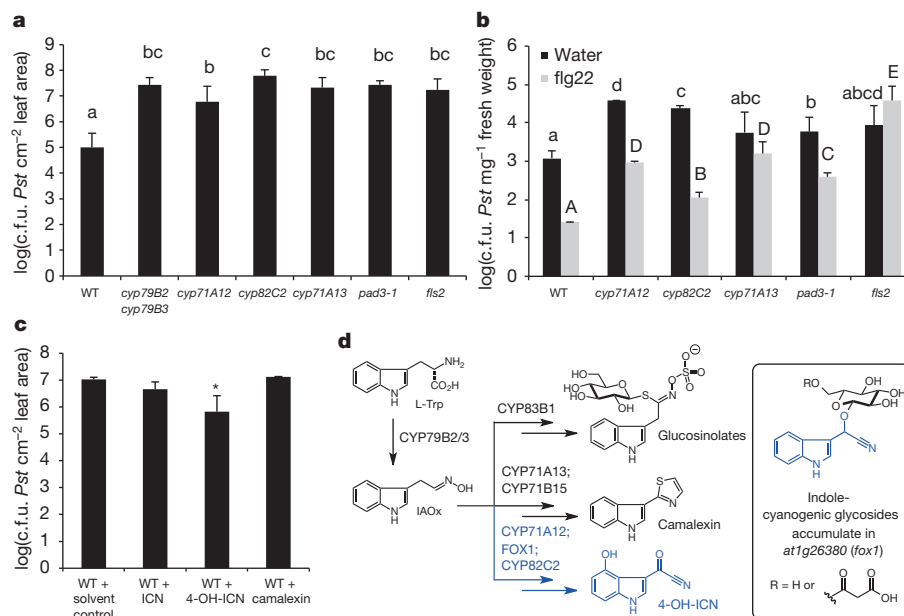
and *cyp71A18* mutants (Fig. 2a). Levels of camalexin and other indolic metabolites were only slightly changed in whole-seedling tissue extracts of the *cyp71A12* mutant (Extended Data Fig. 5c).

Further correlation analysis using *CYP71A12* as bait revealed a cluster of five tandemly arrayed homologous genes, *At1g26380–At1g26420*, that are highly coexpressed with *CYP71A12* (Supplementary Table 3). *At1g26380* encodes a flavin-dependent oxidoreductase known as *FOX1* (ref. 19). We profiled the corresponding homozygous transfer-DNA insertion lines for these genes and found a three- to fivefold reduction in levels of ICN metabolites in the *fox1* mutant, with no significant changes observed for the other mutants (Fig. 2a). Additionally, we observed a build-up of IAL, the expected hydrolysis

product of the indole-3-cyanohydrin intermediate (Extended Data Fig. 5d). More strikingly, the *fox1* mutant accumulates new mass signals corresponding to indole cyanogenic glycosides (ICGs), not previously observed in plants (Extended Data Fig. 6a–e, structures shown in Fig. 4d). Cyanogenic glycoside compounds are widely distributed in the plant kingdom, but have not yet been detected in *Arabidopsis*<sup>9</sup>. Disruption of the ICN pathway at the *FOX1*-catalysed step therefore leads to capture of some portion of the cyanohydrin intermediate by non-specific glycosyltransferases, exactly paralleling dhurrin synthesis<sup>8</sup>.

We sought to confirm the proposed biochemical transformations (Fig. 2c) by reconstituting the complete pathway *in vitro*. A combination of yeast microsomal *CYP71A12* and *CYP82C2* and *N. benthamiana*-expressed *FOX1* was sufficient to catalyse the conversion of IAOx to ICN, as illustrated in Fig. 3; the production of 4-OH-ICN is inferred from the accumulation of 4-OH-ICA. We also reconstituted the biosynthesis of 4-OH-ICN in the heterologous host *N. benthamiana*, using transient expression of the four pathway genes necessary for production of 4-OH-ICN from Trp via *Agrobacterium*-mediated transient transformation<sup>20</sup>. We observed significant accumulation of **B1** (from methanol extraction of 4-OH-ICN) only when all pathway genes were present; however, we also noted background levels of ICA and IAL when only early pathway genes were expressed (Extended Data Fig. 7). Notably, when we expressed *CYP79B2* and *CYP71A12* but not *FOX1*, we again observed the accumulation of ICG mass signals (Extended Data Fig. 6f).

The Trp-derived metabolites camalexin and 4-methoxy indol-3-ylmethylglucosinolate (4-methoxyglucobrassicin) have been shown to play a key role in *Arabidopsis* immunity (Fig. 4d)<sup>10,11,18,21</sup>. To evaluate whether 4-OH-ICN pathway products also contribute to *Arabidopsis* disease resistance, we challenged 4-OH-ICN biosynthetic mutants with a diverse panel of pathogens. Using surface inoculation to mimic the natural infection process, we found that, compared with



**Figure 4 | Camalexin and CYP82C2-synthesized 4-OH-ICN contribute non-redundantly to disease resistance against the virulent bacterial pathogen *P. syringae*.** **a**, Growth analysis of the virulent *P. syringae* pv. *tomato* DC3000 (*Pst*) in surface-inoculated adult leaves. Data represent the mean  $\pm$  s.e.m. of four biological replicates. Data points labelled with different letters are significantly different ( $P < 0.05$ , two-tailed *t* test); data points labelled with the same letter are not significantly different. WT, Col-0 ecotype; c.f.u., colony-forming units. **b**, Growth analysis of *Pst* in 10-day-old seedlings pre-treated with water or 1  $\mu$ M bacterial MAMP flg22 for 6 h. Data represent the median  $\pm$  s.e.m. of four biological replicates of 10–15 seedlings each. Different

letters denote statistically significant differences ( $P < 0.05$ , two-tailed *t* test).

**c**, Growth analysis of *Pst* in WT adult leaves pre-immunized with 1  $\mu$ M flg22 and 100  $\mu$ M ICN, 4-OH-ICN, camalexin or solvent control (dimethylsulfoxide (DMSO)) for 24 h before infiltration with *Pst*. Data represent the median  $\pm$  s.e.m. of three biological replicates. Asterisk denotes statistical significance relative to WT ( $P < 0.01$ , two-tailed *t* test). Experiment was repeated three times, producing similar results. **d**, Summary of known major Trp-derived secondary metabolites in *Arabidopsis* and oxidative biosynthetic enzymes that have been used to reconstitute the pathways *in vitro* or *in planta*.

WT, the adult leaves of *cyp71A12* and *cyp82C2* are more susceptible to the virulent bacterial hemibiotroph *Pst* (*P. syringae* pv. *tomato* DC3000) and comparable to the immuno-deficient *fls2* mutant, which cannot perceive the bacterial microbe-associated molecular pattern (MAMP) flg22 (refs 22, 23) (Fig. 4a). Similarly, seedlings of the 4-OH-ICN pathway mutants are more susceptible to *Pst* than WT in the presence and absence of flg22 (Fig. 4b), indicating a role for 4-OH-ICN in basal disease resistance against a bacterial pathogen. Notably, the adult leaves and seedlings of the camalexin pathway mutants *cyp71A13* and *pad3* are also more susceptible to *Pst* infection than WT (Fig. 4a, b), suggesting a previously unrecognized role for camalexin in the antibacterial defence response. To test for a direct role of the ICN pathway metabolites in the plant innate immune response, either as inducible antibacterial or signalling compounds, we measured their protective effect against subsequent bacterial infection by infecting WT adult leaves with *Pst* after pre-immunizing them with pure compounds and flg22. Compared with a solvent control, pre-treatment with 4-OH-ICN (but not ICN or camalexin) conferred greater bacterial resistance (Fig. 4c), which supports a direct mechanism of action for 4-OH-ICN in inducible plant defence.

We also observed increased disease symptoms in adult leaves of the *cyp82C2* mutant upon inoculation with spores from the avirulent fungal necrotroph *Alternaria brassicicola* (Extended Data Fig. 8e, f) and—consistent with a previous report<sup>24</sup>—the virulent necrotroph *Botrytis cinerea* (Extended Data Fig. 8a, b), but not from the obligate fungal biotroph *Golovinomyces orontii* (Extended Data Fig. 8c, d). Furthermore, purified ICN and 4-OH-ICN have a growth inhibitory effect on *B. cinerea* and *A. brassicicola* comparable to that of camalexin<sup>25</sup> (Extended Data Fig. 9). However, we cannot rule out the possibility that the role of the 4-OH-ICN pathway in fungal defence is indirect, as adult leaves of the *cyp82C2* mutant appear partly impaired in camalexin production after *Alternaria* treatment (Extended Data Fig. 10).

The camalexin and 4-OH-ICN pathways rely on a pair of paralogous genes, *CYP71A12* and *CYP71A13*, which are members of the CYP71 family linked to innovations in plant metabolism<sup>26</sup> (Fig. 4d). Strikingly, the 4-OH-ICN pathway resembles the widespread cyanogenic glucoside pathway that has been lost in the Brassicaceae, and appears to be a metabolic re-invention leading to a novel cyanogenic metabolite type derived from Trp<sup>27,28</sup>. It is possible that 4-OH-ICN acts in concert with other Trp-derived metabolites, each contributing to protection against overlapping sets of specific pathogens. Collectively, our data provide additional insight into the *Arabidopsis* defence response and, more generally, how plants use metabolic innovation to expand innate immunity.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 14 August 2014; accepted 14 July 2015.**

**Published online 9 September 2015.**

- Chae, L., Kim, T., Nilo-Poyanco, R. & Rhee, S. Y. Genomic signatures of specialized metabolism in plants. *Science* **344**, 510–513 (2014).
- D'Auria, J. C. & Gershenzon, J. The secondary metabolism of *Arabidopsis thaliana*: growing like a weed. *Curr. Opin. Plant Biol.* **8**, 308–316 (2005).
- Bednarek, P. & Osbourn, A. Plant-microbe interactions: chemical diversity in plant defense. *Science* **324**, 746–748 (2009).
- Denoux, C. et al. Activation of defense response pathways by OGs and Flg22 elicitors in *Arabidopsis* seedlings. *Mol. Plant* **1**, 423–445 (2008).

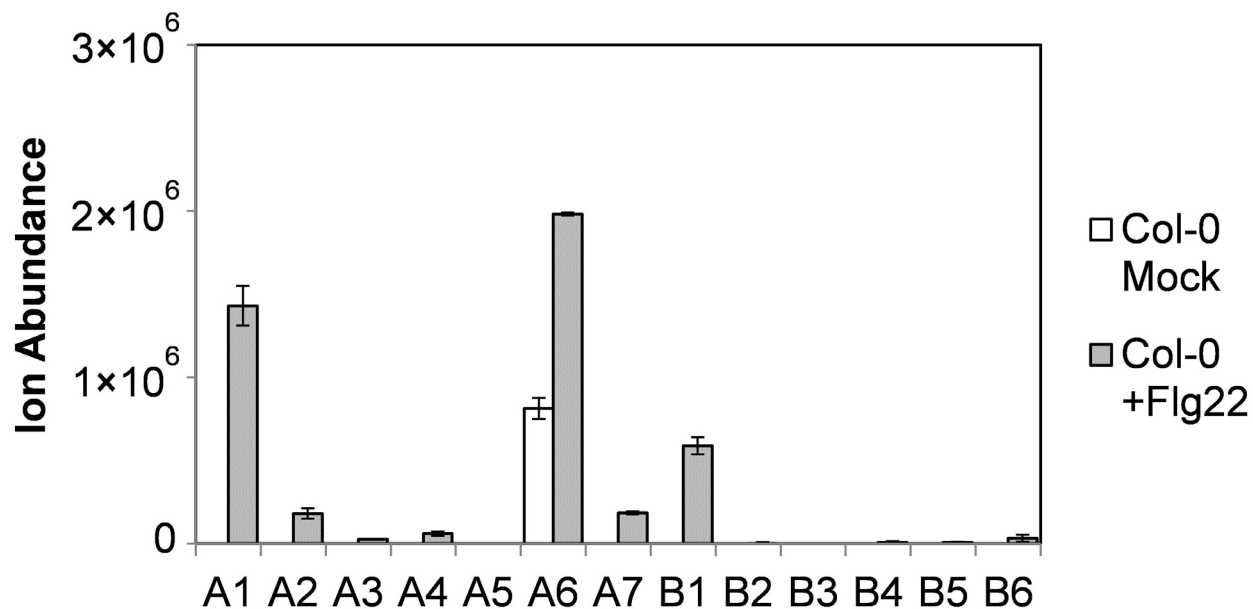
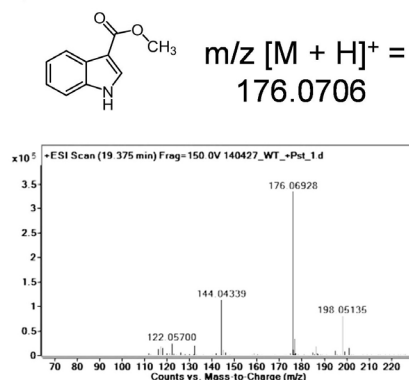
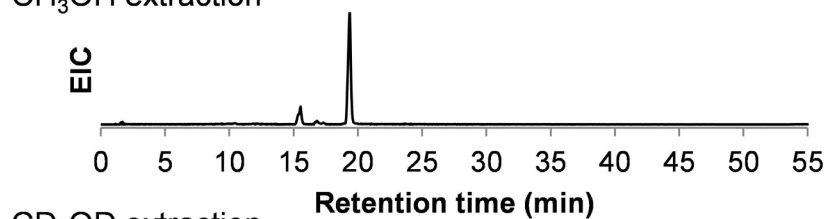
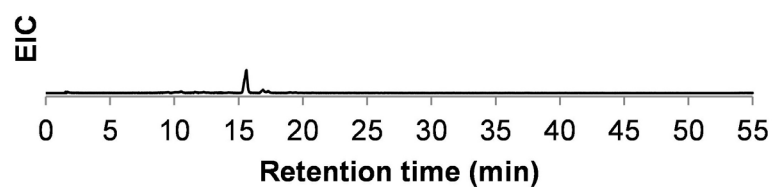
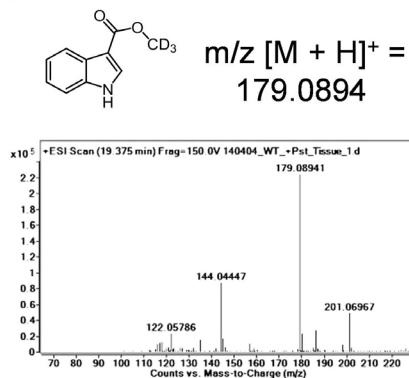
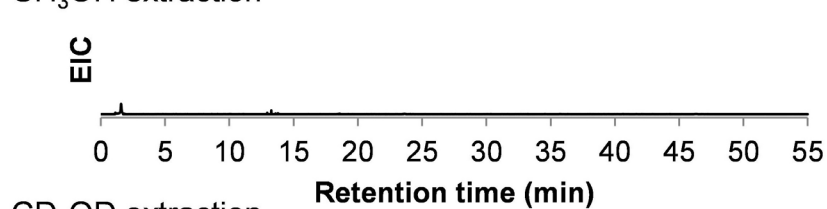
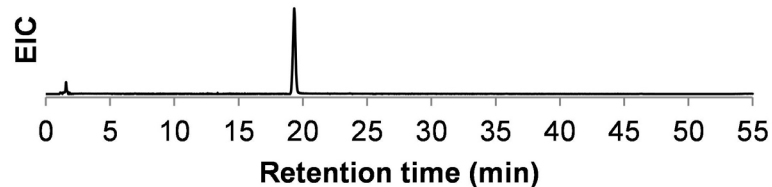
- Bak, S. et al. Cytochromes P450. *Arabidopsis Book* **9**, e0144 (2011).
- Jones, T. H., Conner, W. E., Meinwald, J., Eisner, H. E. & Eisner, T. Benzoyl cyanide and mandelonitrile in the cyanogenic secretion of a centipede. *J. Chem. Ecol.* **2**, 421–429 (1976).
- Zagrobelyny, M., Bak, S. & Møller, B. L. Cyanogenesis in plants and arthropods. *Phytochemistry* **69**, 1457–1468 (2008).
- Gleadow, R. M. & Møller, B. L. Cyanogenic glycosides: synthesis, physiology, and phenotypic plasticity. *Annu. Rev. Plant Biol.* **65**, 155–185 (2014).
- Tattersall, D. B. et al. Resistance to an herbivore through engineered cyanogenic glucoside synthesis. *Science* **293**, 1826–1828 (2001).
- Bednarek, P. et al. A glucosinolate metabolism pathway in living plant cells mediates broad-spectrum antifungal defense. *Science* **323**, 101–106 (2009).
- Clay, N. K., Adio, A. M., Denoux, C., Jander, G. & Ausubel, F. M. Glucosinolate metabolites required for an *Arabidopsis* innate immune response. *Science* **323**, 95–101 (2009).
- Dangl, J. L., Horvath, D. M. & Staskawicz, B. J. Pivoting the plant immune system from dissection to deployment. *Science* **341**, 746–751 (2013).
- Ahuja, I., Kissen, R. & Bones, A. M. Phytoalexins in defense against pathogens. *Trends Plant Sci.* **17**, 73–90 (2012).
- Vinayavekhin, N. & Saghatelian, A. In *Current Protocols in Molecular Biology* (eds Ausubel, F. M. et al.) Ch. 30 (2010).
- Böttcher, C. et al. The biosynthetic pathway of indole-3-carbaldehyde and indole-3-carboxylic acid derivatives in *Arabidopsis*. *Plant Physiol.* **165**, 841–853 (2014).
- Mikkelsen, M. D., Hansen, C. H., Wittstock, U. & Halkier, B. A. Cytochrome P450 CYP79B2 from *Arabidopsis* catalyzes the conversion of tryptophan to indole-3-acetaldoxime, a precursor of indole glucosinolates and indole-3-acetic acid. *J. Biol. Chem.* **275**, 33712–33717 (2000).
- Zhao, Y. et al. Trp-dependent auxin biosynthesis in *Arabidopsis*: involvement of cytochrome P450s CYP79B2 and CYP79B3. *Genes Dev.* **16**, 3100–3112 (2002).
- Millet, Y. A. et al. Innate immune responses activated in *Arabidopsis* roots by microbe-associated molecular patterns. *Plant Cell* **22**, 973–990 (2010).
- Boudsocq, M. et al. Differential innate immune signalling via Ca<sup>2+</sup> sensor protein kinases. *Nature* **464**, 418–422 (2010).
- Peyret, H. & Lomonosoff, G. P. The pEAQ vector series: the easy and quick way to produce recombinant proteins in plants. *Plant Mol. Biol.* **83**, 51–58 (2013).
- Thomma, B. P., Nelissen, I., Eggermont, K. & Broekaert, W. F. Deficiency in phytoalexin production causes enhanced susceptibility of *Arabidopsis thaliana* to the fungus *Alternaria brassicicola*. *Plant J.* **19**, 163–171 (1999).
- Gómez-Gómez, L. & Boller, T. Flagellin perception: a paradigm for innate immunity. *Trends Plant Sci.* **7**, 251–256 (2002).
- Zipfel, C. et al. Bacterial disease resistance in *Arabidopsis* through flagellin perception. *Nature* **428**, 764–767 (2004).
- Liu, F. et al. The *Arabidopsis* P450 protein CYP82C2 modulates jasmonate-induced root growth inhibition, defense gene expression and indole glucosinolate biosynthesis. *Cell Res.* **20**, 539–552 (2010).
- Nafisi, M. et al. *Arabidopsis* cytochrome P450 monooxygenase 71A13 catalyzes the conversion of indole-3-acetaldoxime in camalexin synthesis. *Plant Cell* **19**, 2039–2052 (2007).
- Nelson, D. & Werck-Reichhart, D. A P450-centric view of plant evolution. *Plant J.* **66**, 194–211 (2011).
- Møller, B. L. Functional diversifications of cyanogenic glucosides. *Curr. Opin. Plant Biol.* **13**, 338–347 (2010).
- Rauhut, T. & Glawischning, E. Evolution of camalexin and structurally related indolic compounds. *Phytochemistry* **70**, 1638–1644 (2009).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank F. Ausubel, M. B. Mudgett, C. Khosla, A. Saghatelian, Y. Millet, C. Danna, S. Galanie, and members of the Sattely and Clay laboratories for advice on experiments and comments on the manuscript. We thank the Salk Institute Genomic Analysis Laboratory for providing the sequence-indexed *Arabidopsis* transfer-DNA insertion mutants. We thank G. Lomonosoff (John Innes Centre) for providing plasmid pEAQ. This work was supported by R00 GM089985 and DP2 AT008321 (to E.S.S.), T32 GM008412-20 (to J.R.), and T32 GM007499-38 (to B.B.). The early stages of this work were supported by National Science Foundation grant MCB-0519898 and National Institutes of Health grant R37 GM 48707 (awarded to Fred Ausubel, Massachusetts General Hospital, Boston, Massachusetts, USA).

**Author Contributions** J.R., B.B., N.K.C., and E.S.S. designed experiments. J.R. and B.B. performed experiments. J.R., B.B., N.K.C., and E.S.S. analysed data and wrote the paper.

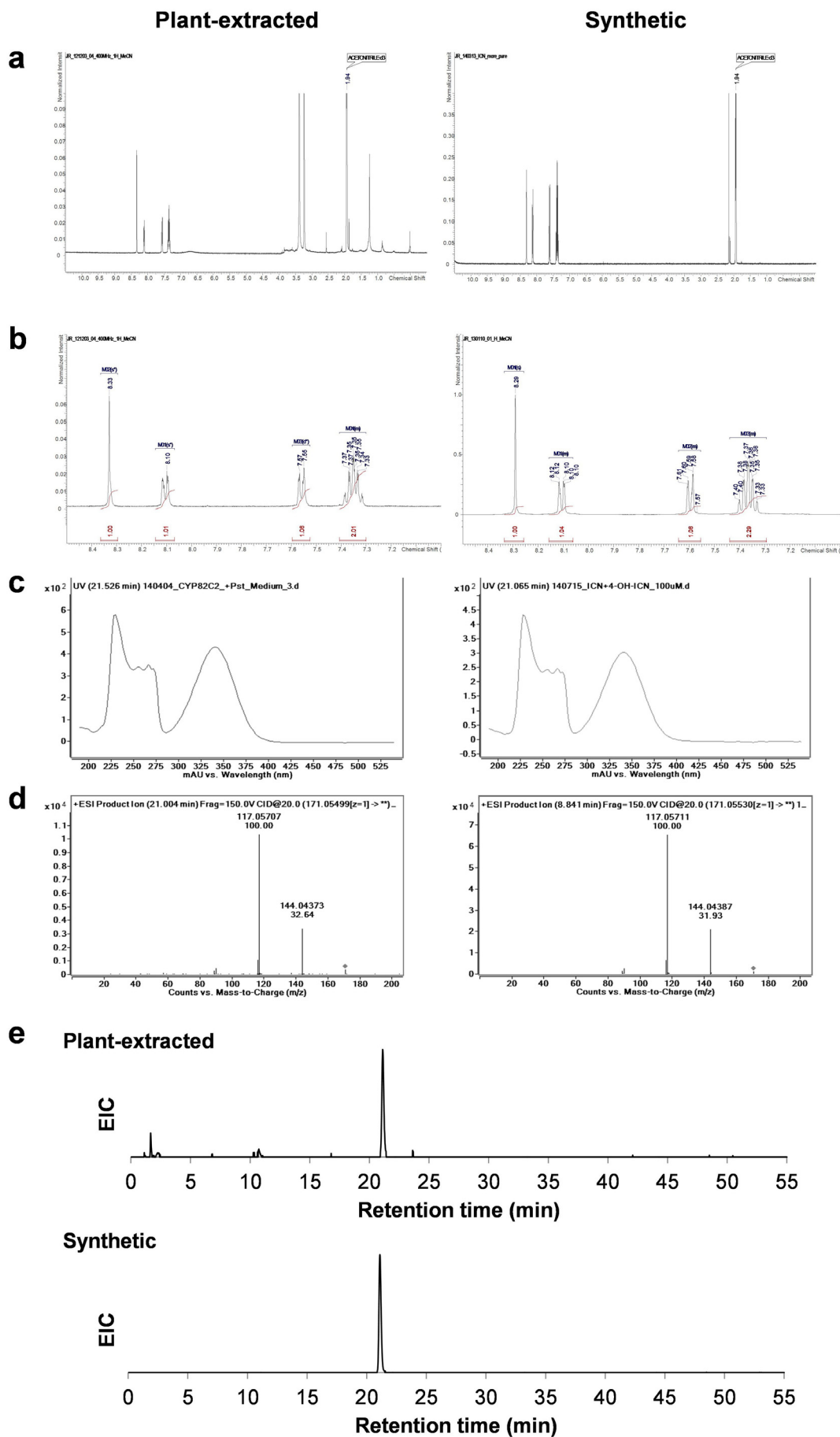
**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to E.S.S. ([sattely@stanford.edu](mailto:sattely@stanford.edu)) or N.K.C. ([nicole.clay@yale.edu](mailto:nicole.clay@yale.edu)).

**a****b****c** CH<sub>3</sub>OH extractionCD<sub>3</sub>OD extraction**d****e** CH<sub>3</sub>OH extractionCD<sub>3</sub>OD extraction

**Extended Data Figure 1 | Elicitation of compounds identified in metabolomics screen by flg22 peptide and origin of ICA methyl ester as artefact of the methanol extraction method.** **a**, Levels of compounds in Flg22-elicited *Arabidopsis* Col-0 seedling tissue, quantified as mean  $[M + H]^+$  ion ( $m/z \pm 10$  ppm) abundances extracted from raw data; error bars, s.d. based on three biological replicates. Production of these compounds in axenic plant culture demonstrates that they are plant-derived. **b, c**, Structure and mass peaks of ICA methyl ester (compound **A1**) seen in LC-MS analysis (**b**), and EICs for the expected  $m/z$  using a standard extraction with 80:20

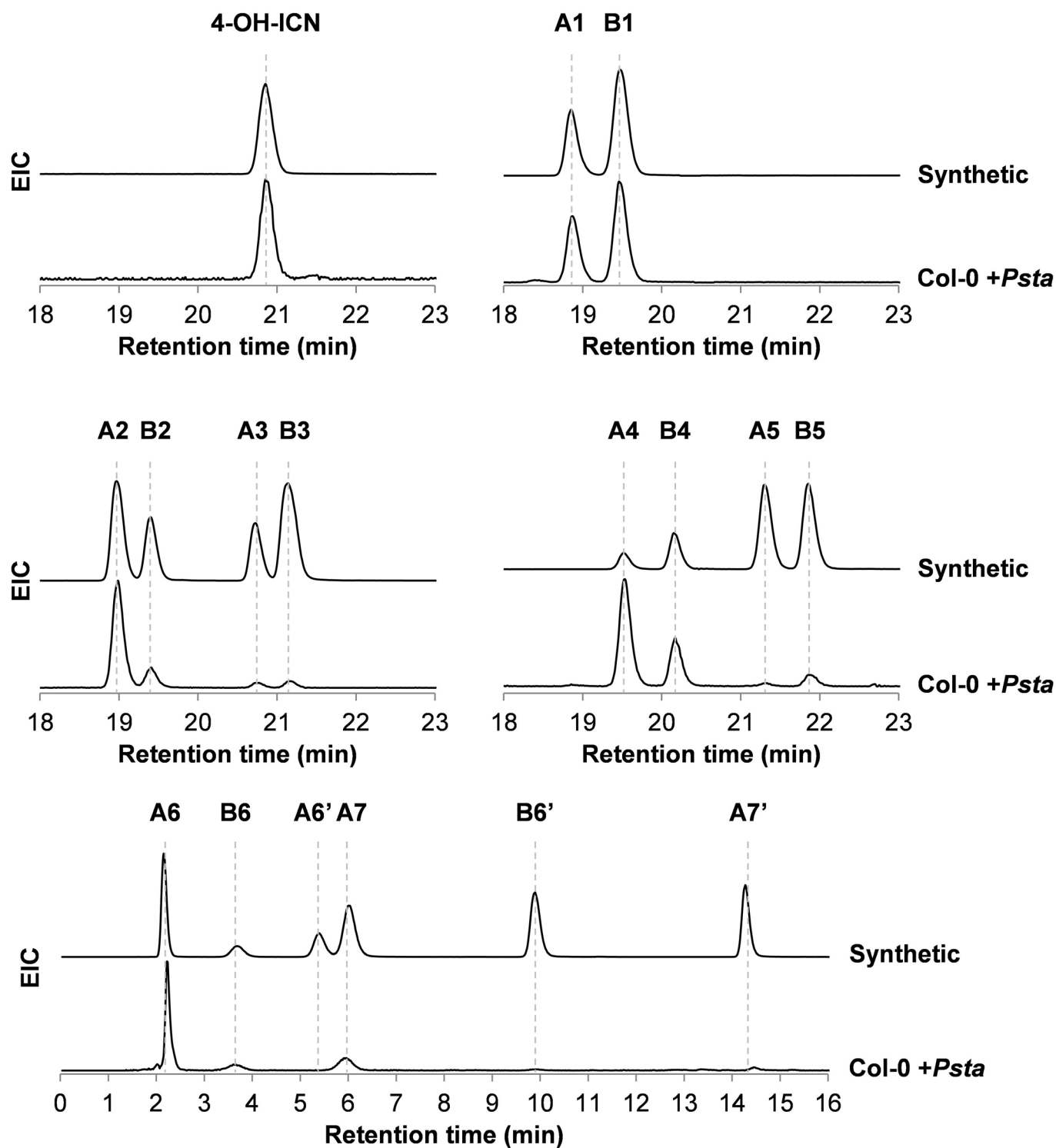
$\text{CH}_3\text{OH}/\text{H}_2\text{O}$  or with 80:20  $\text{CD}_3\text{OD}/\text{D}_2\text{O}$  (**c**). **d, e**, Structure and mass spectrum peaks seen for the triply deuterated **A1** analogue (**d**), and EICs for the expected  $m/z$  using extraction with 80:20  $\text{CH}_3\text{OH}/\text{H}_2\text{O}$ , or with 80:20  $\text{CD}_3\text{OD}/\text{D}_2\text{O}$  (all EICs are to scale) (**e**). The presence of the deuterated analogue of ICA methyl ester and the complete absence of the non-deuterated compound in plant extracts when  $\text{CD}_3\text{OD}$  is substituted for  $\text{CH}_3\text{OH}$  show that the methyl ester is not a product of *Arabidopsis* metabolism, but arises because of the extraction method as a degradation product of ICN.





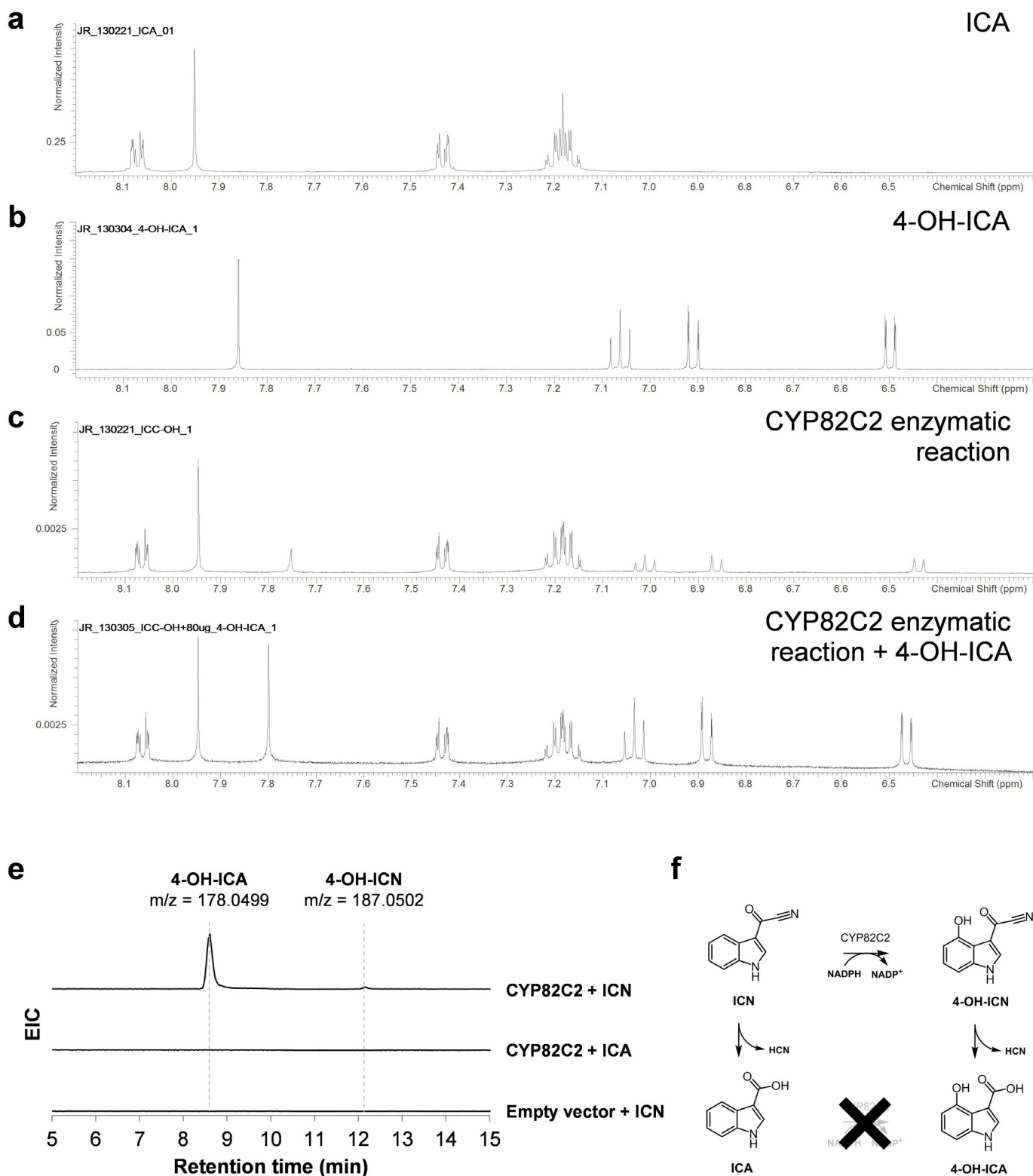
**Extended Data Figure 2 | Comparison of spectra for plant-extracted and synthetic compound establishes identity of ICN as new indolic metabolite produced by *A. thaliana*.** **a**, Full-range ( $\delta = 10.5$  to  $-0.5$ ) and, **b**, downfield region partial ( $\delta = 8.5$ – $7.0$ )  $^1\text{H}$  NMR spectra in  $\text{CD}_3\text{CN}$ . Upfield contaminants in the full-range spectra are presumed to be residual solvent. **c**, Ultraviolet–visible absorbance spectra obtained via a diode array detector during LC

analysis. Note that the prominent peak at 230 nm is due to acetonitrile in the LC mobile phase. **d**, Targeted MS/MS spectra for the parent ICN  $[\text{M} + \text{H}]^+$  ion ( $m/z = 171.0550$ ) at a collision energy of 20 V. See Supplementary Table 1 for relative peak intensities at other collision energies. **e**, Aligned EICs for the ICN  $[\text{M} + \text{H}]^+$  ion for a Col-0 + *Psta* tissue sample extracted with DMSO and synthetic compound, showing identical retention times.



**Extended Data Figure 3** | Comparison of plant-extracted ICN derivatives, 4-OH-ICN derivatives, and synthetic standards shows identical column elution times for all compounds. Col-0 + *Psta* combined EICs were extracted for the relevant compound  $[M + H]^+$   $m/z$  values for a DMSO-extracted medium sample (4-OH-ICN trace), or a MeOH-extracted seedling tissue

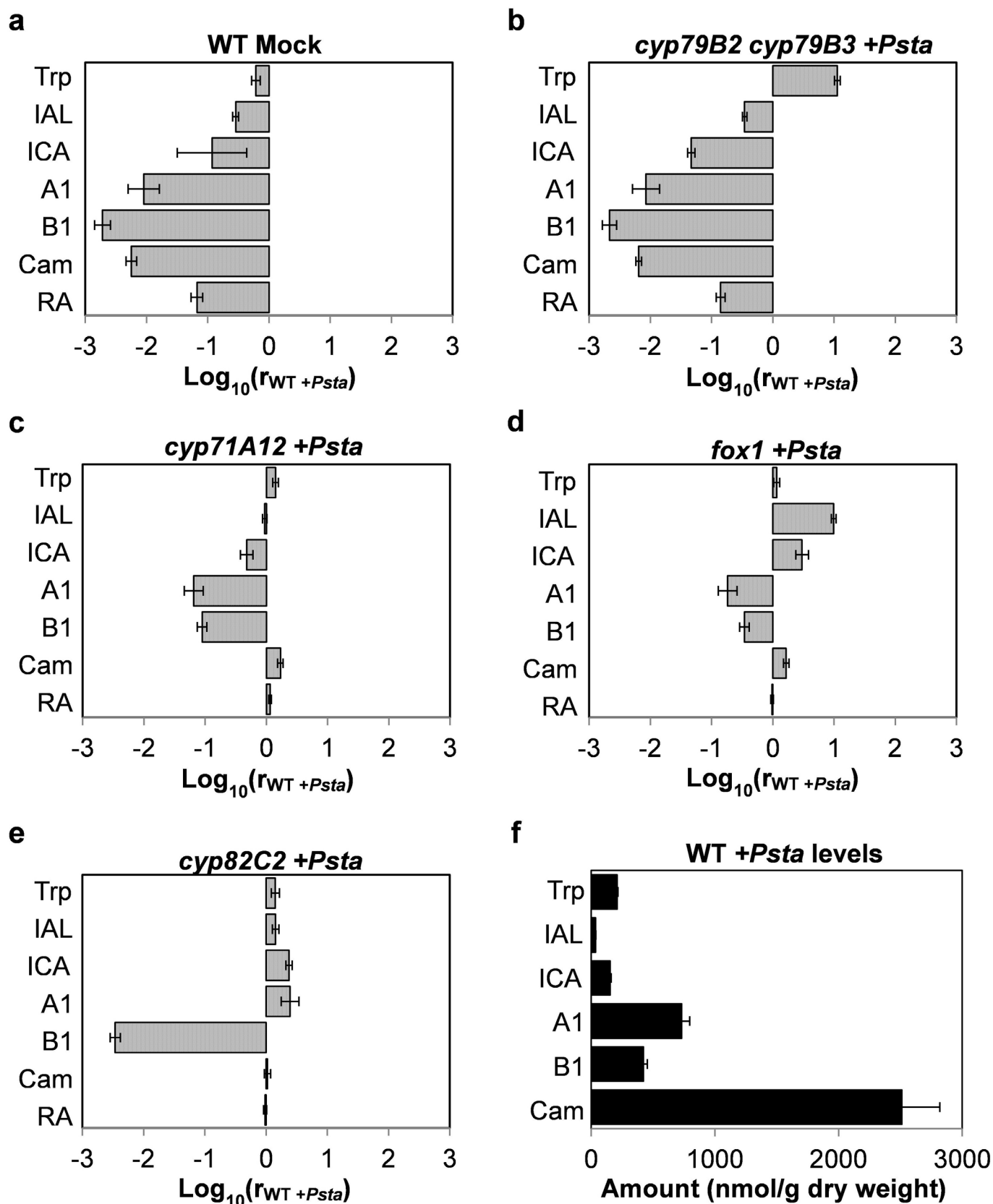
sample (all other traces), while synthetic EICs were extracted for a mixed standard in DMSO. Note that chromatograms are not to scale, and the synthetic standard is not equimolar with respect to all compounds because of partial degradation.



**Extended Data Figure 4 | CYP82C2 is an ICN 4-hydroxylase.** **a, b,**  $^1\text{H}$  NMR spectra in  $\text{CD}_3\text{OD}$  of synthetic ICA (**a**) and 4-OH-ICA (**b**). **c,** Spectrum for large-scale enzymatic reaction extract of ICN incubated with CYP82C2. In addition to ICA, resulting from hydrolysis of ICN, peaks for a singly hydroxylated analogue of ICA are seen; these are qualitatively consistent with, but shifted slightly upfield ( $\sim 30$ – $60$  Hz) from the 4-OH-ICA spectrum, possibly because of impurities or a pH effect in the enzymatic reaction sample. **d,** To confirm the identity conclusively,  $80\text{ }\mu\text{g}$  of 4-OH-ICA dissolved in  $\text{CD}_3\text{OD}$  was added to the enzymatic reaction NMR sample before acquiring another spectrum: no new peaks are seen, while the prior hydroxylated ICA

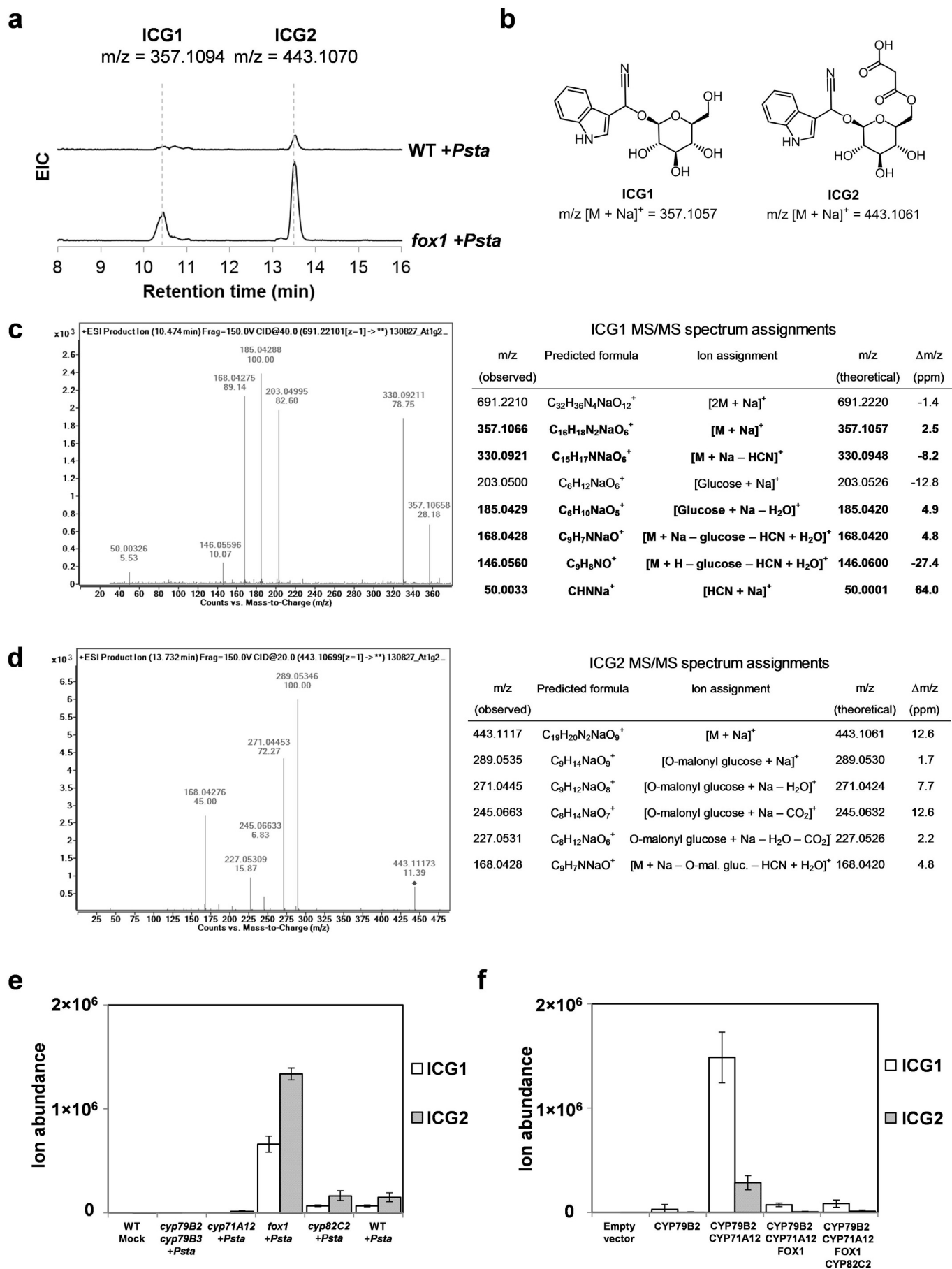
peaks grow in intensity, establishing the product of the enzymatic reaction as 4-OH-ICA. **e**, EICs for enzymatic reactions of CYP82C2 on ICN or ICA, or empty vector control incubation with ICN. Only trace amounts of the expected 4-OH-ICN product but significant amounts of 4-OH-ICA are seen for the CYP82C2/ICN reaction. No hydroxylated products are seen for the CYP82C2/ICA or empty vector/ICN reactions, indicating that CYP82C2 catalyses only the hydroxylation of ICN to 4-OH-ICN, but 4-OH-ICA is seen as the predominant end product due to rapid hydrolysis of 4-OH-ICN (**f**). Chromatograms in this figure were obtained using the 20 min LC-MS gradient (see Supplementary Information, Methods section 1.9 LC-MS analysis).





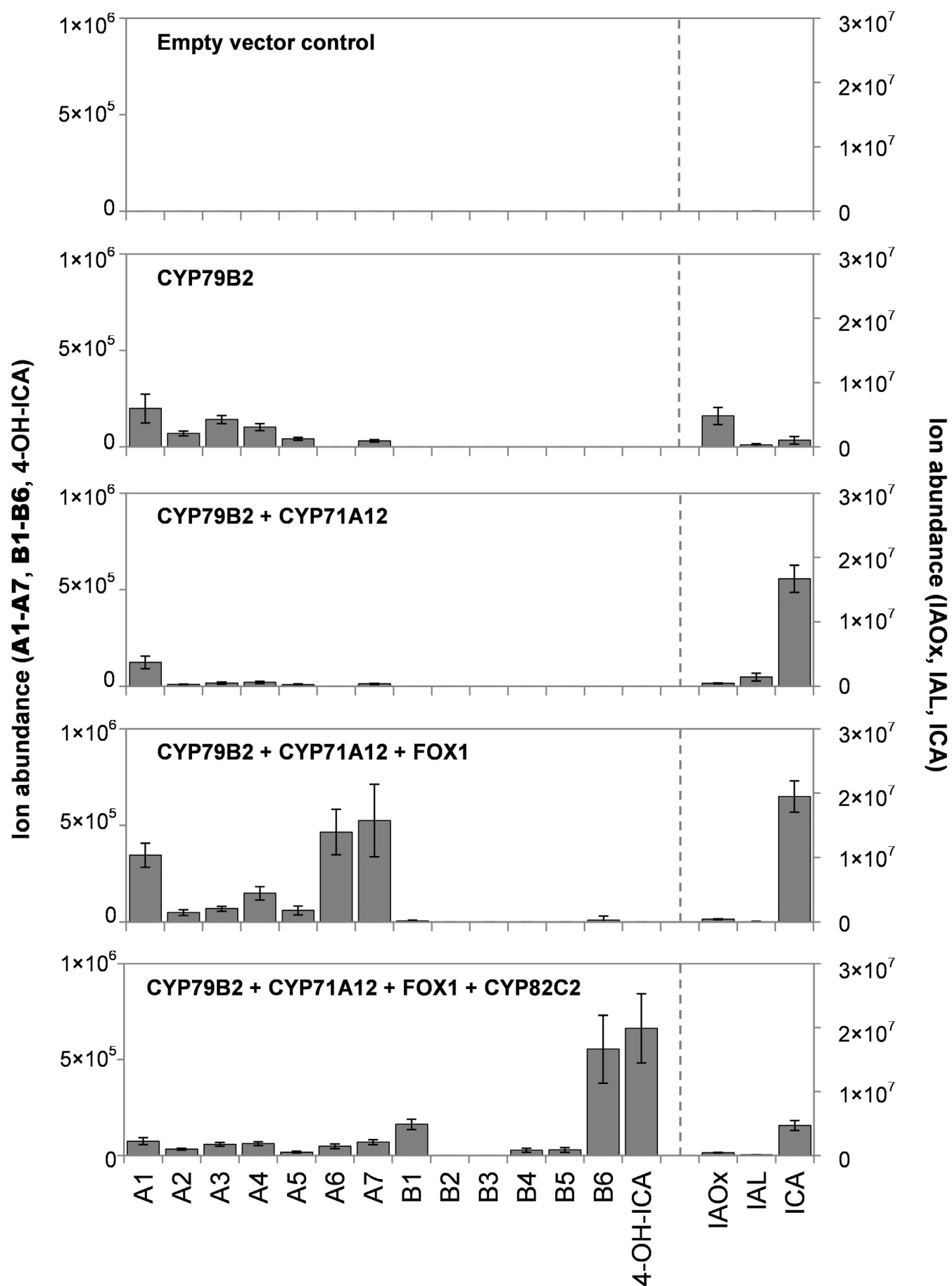
**Extended Data Figure 5 | Levels of numerous *Arabidopsis* indolic metabolites are altered in ICN pathway gene insertion lines compared with WT plants.** a–e, Relative compound levels for mock treatment condition and indicated pathway insertion line mutants, and, f, absolute levels in *Psta*-treated WT (Col-0) seedlings. For a–e, data bars represent a logarithmically scaled ratio of mean metabolite levels in the indicated line or treatment

condition, quantified as  $[M + H]^+$  ion abundances by LC–MS analysis with XCMS processing, to levels in *Psta*-treated WT *Arabidopsis* seedlings. In f, absolute levels for all compounds except RA were quantified by measuring  $[M + H]^+$  ion abundances and comparing to standard curves. Error bars, s.d., based on six biological replicates. Cam, camalexin; RA, raphanusamic acid; other abbreviations as detailed previously.



**Extended Data Figure 6 | Putative ICGs observed in *Arabidopsis* and in *N. benthamiana* expressing ICN pathway enzymes.** **a**, EICs for putative ICGs in WT *Arabidopsis* and *fox* mutant elicited with *Psta*. The  $m/z$  values shown are median values calculated by XCMS. **b**, Hypothesized structures and theoretical  $m/z$  values for the two ICGs identified. **c**, MS/MS spectrum for ICG1;  $m/z$  values and relative abundances are shown above each peak. The ion analysed here ( $m/z = 691.2210$ ) represents a  $[2M + Na]^+$  dimer that is significantly more abundant than the  $[M + Na]^+$  ion. Direct analysis of the  $[M + Na]^+$  ion ( $m/z = 357.1057$ ) yielded low abundance spectra that could not be easily analysed. At lower collision energies, the  $[2M + Na]^+$  ion fragments to  $[M + Na]^+$ , but yields a rich spectrum at 40 V, which is shown.

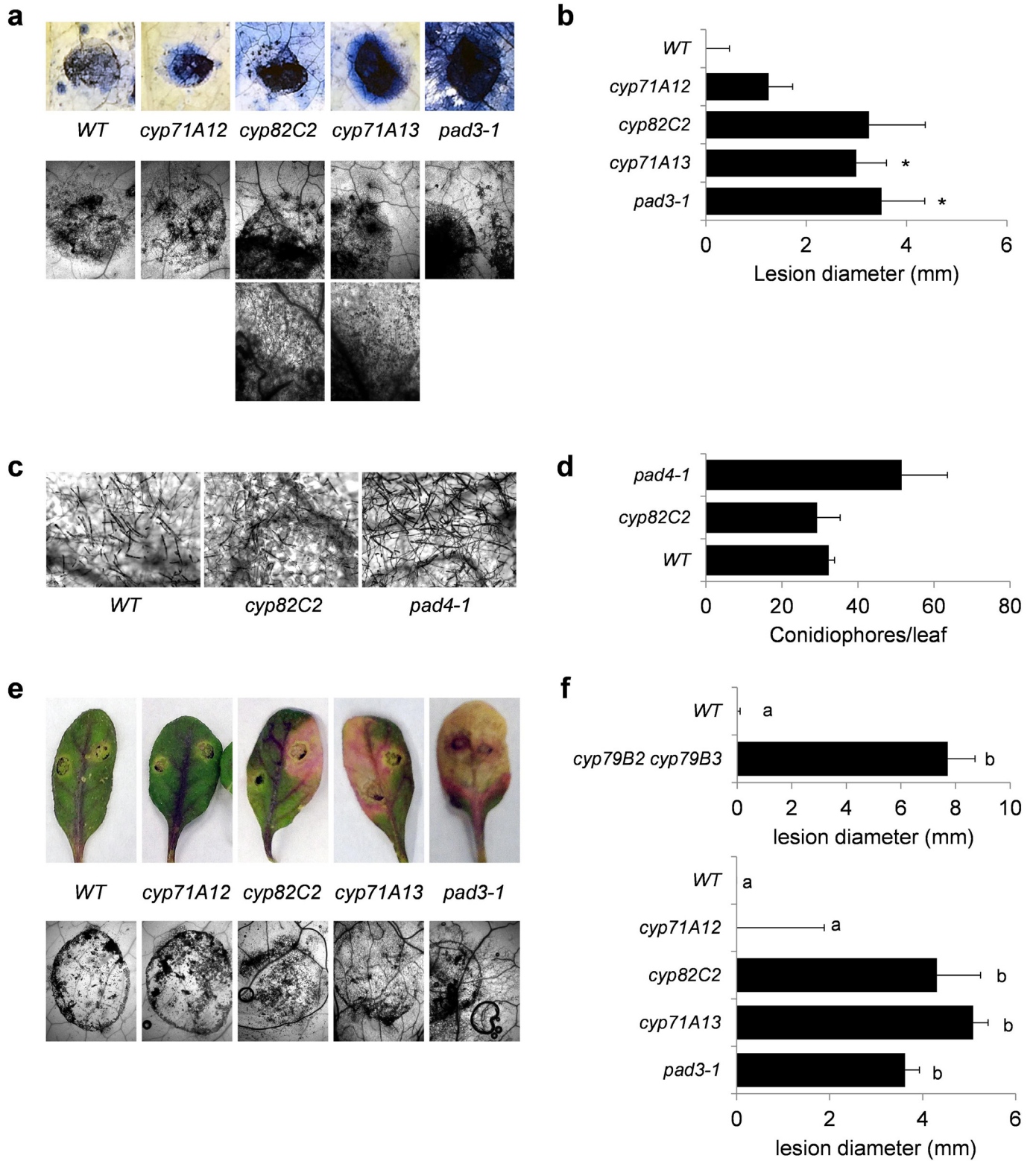
Predicted peak assignments for the ICG1 MS/MS spectrum are shown in the accompanying table. For peaks in bold, exact counterparts could be identified in the dhurrin  $[M + Na]^+$  20 V MS/MS spectrum in the METLIN metabolite database. **d**, MS/MS spectrum obtained for the ICG2  $[M + Na]^+$  ion and predicted peak assignments. While the  $[2M + Na]^+$  peak ( $m/z = 864.2225$ ) is also seen for this compound (not shown),  $[M + Na]^+$  is more abundant in this case, and was analysed directly. **e**, **f**, Levels of ICG1 and ICG2 in ICN pathway mutants (**e**) and in WT plants elicited with *Psta* and *N. benthamiana* expressing ICN pathway enzymes (**f**). For **e** and **f**, levels are quantified as mean  $[M + Na]^+$  ion ( $m/z \pm 10$  ppm) abundances extracted from raw data; error bars, s.d., based on six biological replicates.





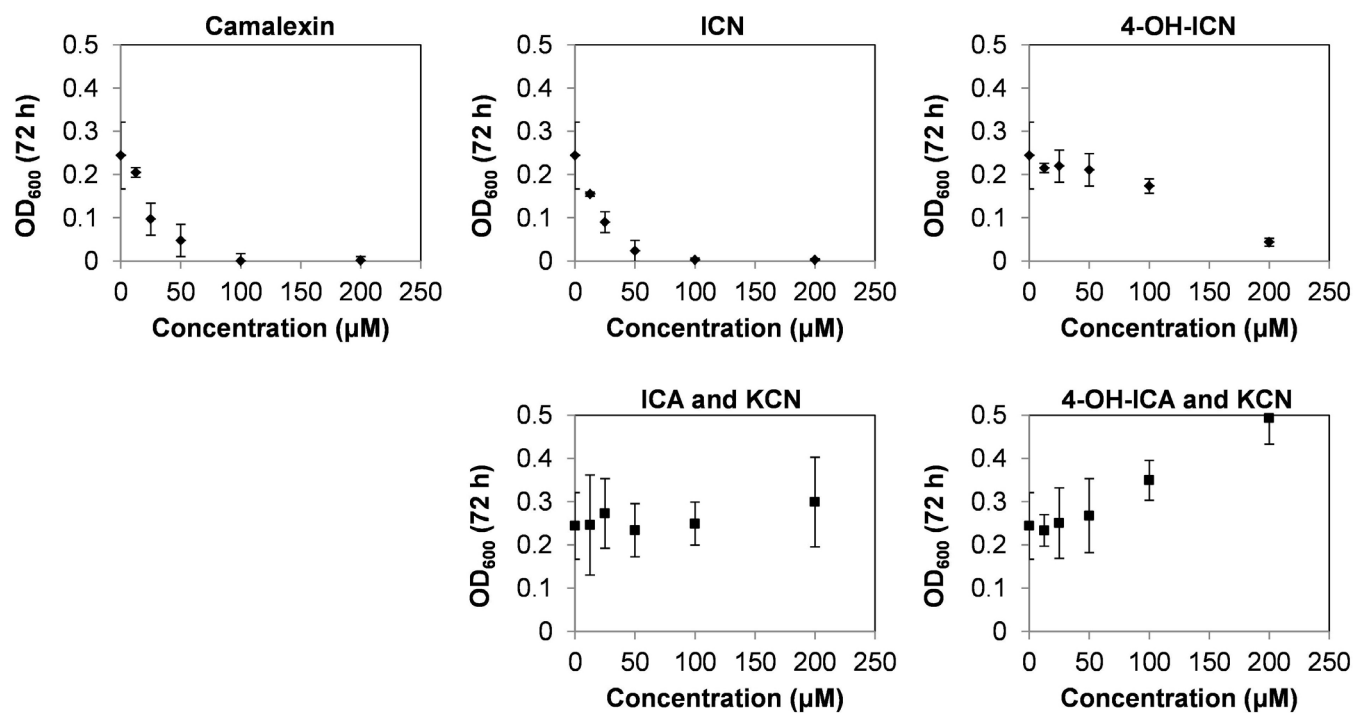
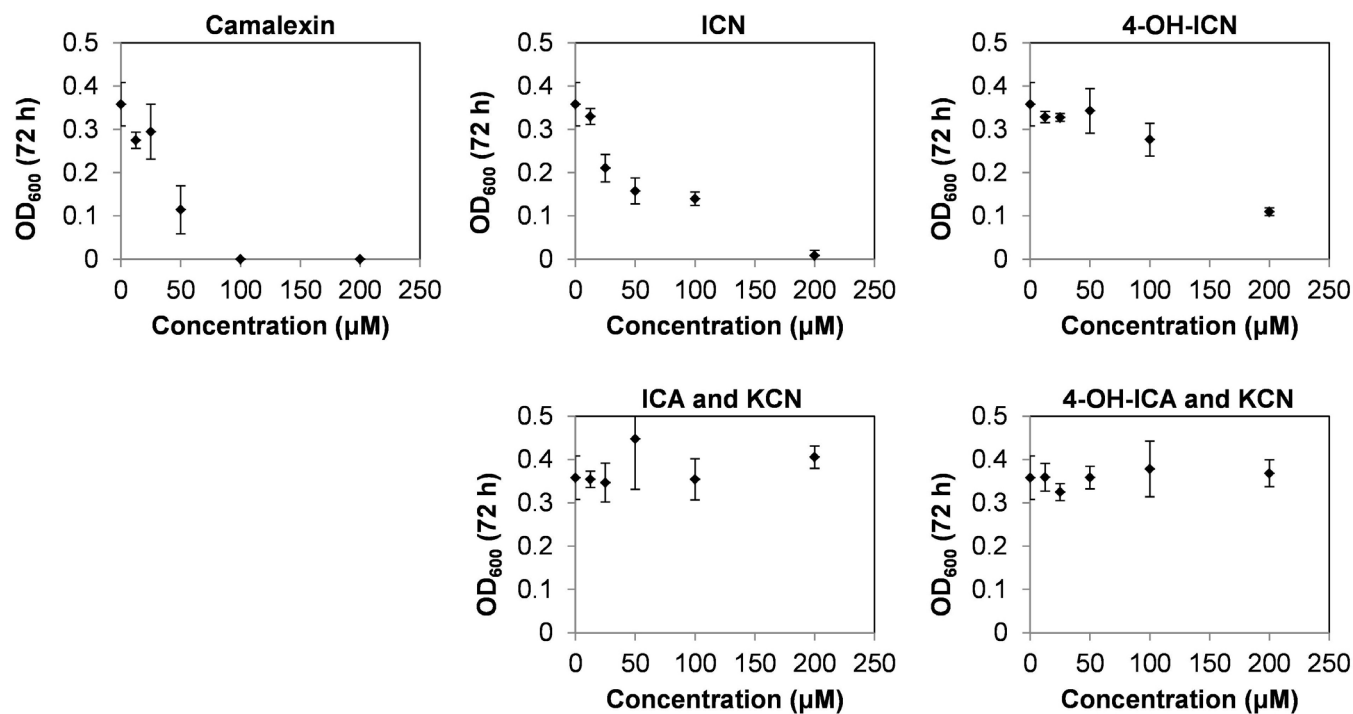
**Extended Data Figure 7 | ICN pathway metabolites are produced in *N. benthamiana* transiently expressing pathway genes.** Levels of ICN and 4-OH-ICN derivatives (left axis) and other relevant indolic compounds (right axis), quantified as mean  $[M + H]^+$  ion ( $m/z \pm 10$  ppm) abundances extracted from raw data; error bars, s.d., based on six biological replicates. The

set of transiently expressed genes is indicated for each panel. Background levels of ICA and IAL detected when only the early pathway genes *CYP71A12* and/or *CYP79B2* are expressed indicate potential involvement of endogenous *N. benthamiana* enzymes.



**Extended Data Figure 8 | ICN pathway metabolites contribute to disease resistance towards *B. cinerea* but not towards *G. orontii*.** **a**, Top: typical lactophenol trypan blue staining of leaves drop-inoculated with spores from the virulent fungal necrotroph *B. cinerea* to visualize the extent of host cell death (darkly stained areas within and beyond the fungal spore droplet region). Middle: microscopic analysis of stained leaves to visualize the extent of fungal colonization (stained filamentous fungal hyphae within and beyond the fungal spore droplet region). Images were taken at the same magnification ( $\times 25$ ) and are representative of five biological replicates. Bottom: close-up images of the fungal hyphae beyond the fungal spore droplet region for *cyp82C2* and *cyp71A13* mutants. Images were taken at the same magnification ( $\times 100$ ). **b**, Measurement of the disease lesion diameters in infected leaves. Data represent the median  $\pm$  s.e.m. for five biological replicates. Asterisks denote statistical significance relative to WT ( $P < 0.05$ , two-tailed  $t$  test). **c**, Typical lactophenol trypan blue staining of fungal conidiophores (spore-bearing structures) formed in leaves infected with the adapted powdery mildew

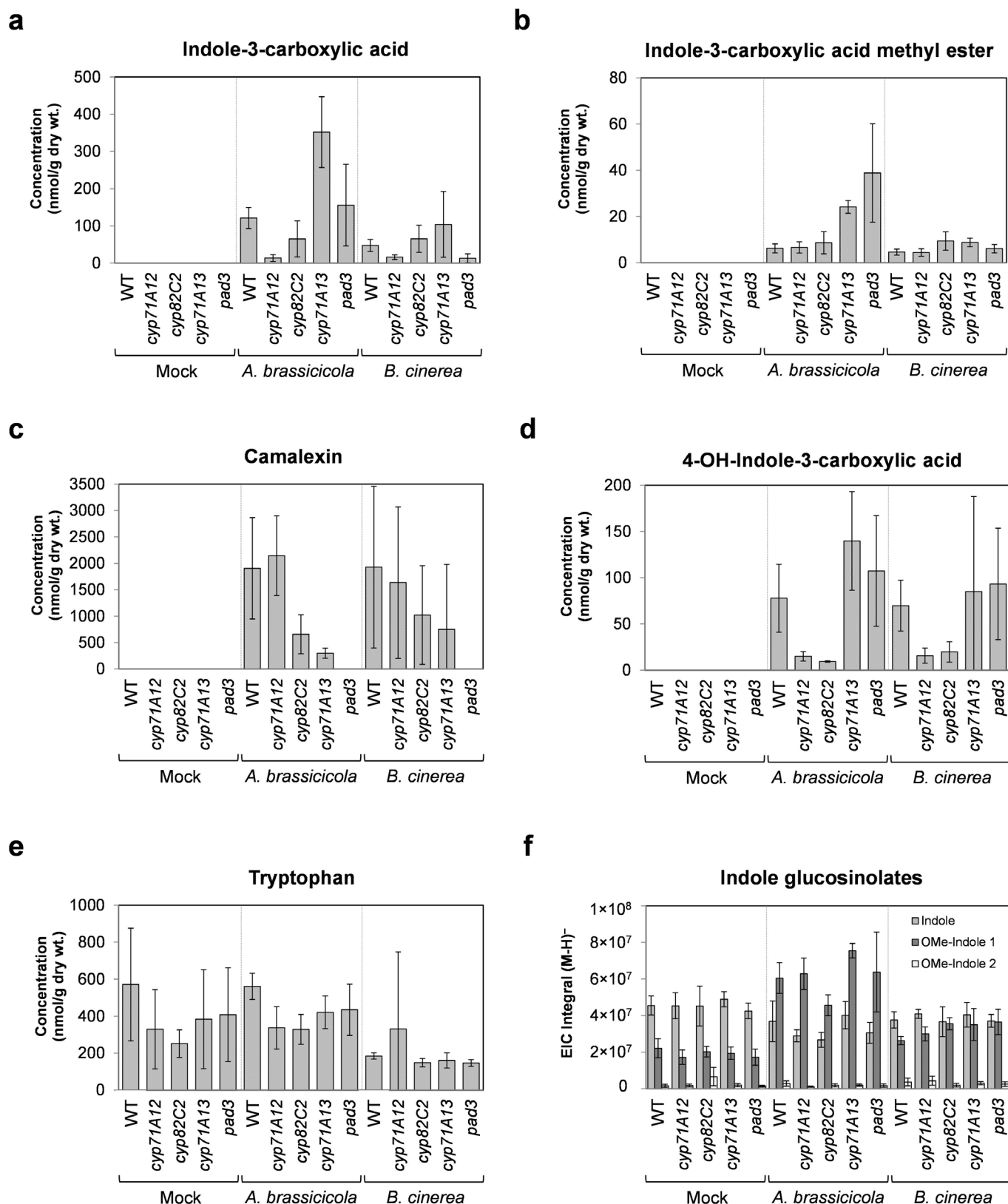
*G. orontii*. The *pad4-1* mutant is more susceptible to fungal growth by *G. orontii* and thus produces significantly more conidiophores. Images were taken at the same magnification ( $\times 100$ ) and are representative of three biological replicates. **d**, Measurement of the number of conidiophores in infected leaves. Data represent the mean  $\pm$  s.d. for three biological replicates. **e**, Top: typical disease symptoms 3 days after drop inoculation of leaves with spores from the avirulent fungal necrotroph *A. brassicicola*. Bottom: microscopic analysis of infected leaves after lactophenol trypan blue staining confirming that disease symptoms are consistent with extent of fungal colonization (lightly stained fungal hyphae extending from the fungal spore droplet region) and host cell death (darkly stained areas along and beyond the border of the spore droplet region). Images were taken at the same magnification ( $\times 25$ ) and are representative of ten biological replicates. **f**, Measurement of the disease lesion diameters in infected leaves. Data represent the median  $\pm$  s.e.m. of eight (top graph) or ten biological replicates (bottom graph). Different letters denote statistically significant differences ( $P < 0.05$ , two-tailed  $t$  test).

**a*****B. cinerea* SF1****b*****A. brassicicola* FSU218**



**Extended Data Figure 9 | ICN and 4-OH-ICN but not their degradation products inhibit fungal growth *in vitro*.** **a, b,** Fungal growth inhibition assays on *B. cinerea* SF1 (**a**) or *A. brassicicola* FSU218 (**b**) with the tested compound (or compound combination) indicated. For compound combinations, the concentration indicated is for each compound; the given combinations approximate the hydrolysis products of ICN or 4-OH-ICN. Growth of fungi in potato dextrose broth on a microplate was quantified by measuring absorbance

at 600 nm ( $OD_{600\text{ nm}}$ ) 72 h after spore inoculation and subtracting the absorbance at 0 h; see Methods for further details. Error bars, s.d. based on three biological replicates. Note that the half-maximum inhibitory concentrations ( $IC_{50}$ ) for both camalexin and ICN are approximately 25  $\mu\text{M}$  against *B. cinerea* and 50  $\mu\text{M}$  against *A. brassicicola*. For 4-OH-ICN, the inhibitory effect is not as pronounced, possibly because of rapid degradation of 4-OH-ICN in potato dextrose broth (see Supplementary Table 2).



**Extended Data Figure 10 | Levels of indolic compounds in leaves of mature plants after mock treatment or fungal infection.** Tissue extracts were analysed by LC-MS 7 days post-infection for *A. brassicicola* FSU218 and 5 days post-infection for *B. cinerea* SF1. **a–e**, Levels of indicated compound, quantified

as EIC integral for the  $[M + H]^+$  ion ( $m/z \pm 10$  ppm) and converted to absolute amounts by comparison with a standard curve. **f**, Ion count integrals for indole glucosinolates ( $[M-H]^-$  ion,  $m/z \pm 10$  ppm). Error bars in all panels, s.d. based on six biological replicates.

# Erosion of the chronic myeloid leukaemia stem cell pool by PPAR $\gamma$ agonists

Stéphane Prost<sup>1</sup>, Francis Relouzat<sup>1</sup>, Marc Spentchian<sup>2</sup>, Yasmine Ouzegdough<sup>1</sup>, Joseph Saliba<sup>1</sup>, Gérald Massonnet<sup>3</sup>, Jean-Paul Beressi<sup>4</sup>, Els Verhoeyen<sup>5,6</sup>, Victoria Raggueneau<sup>7</sup>, Benjamin Maneglier<sup>8</sup>, Sylvie Castaigne<sup>9</sup>, Christine Chomienne<sup>3</sup>, Stany Chrétien<sup>1,10\*</sup>, Philippe Rousselot<sup>3,9\*</sup> & Philippe Leboulch<sup>1,11,12\*</sup>

Whether cancer is maintained by a small number of stem cells or is composed of proliferating cells with approximate phenotypic equivalency is a central question in cancer biology<sup>1</sup>. In the stem cell hypothesis, relapse after treatment may occur by failure to eradicate cancer stem cells. Chronic myeloid leukaemia (CML) is quintessential to this hypothesis. CML is a myeloproliferative disorder that results from dysregulated tyrosine kinase activity of the fusion oncoprotein BCR-ABL<sup>2</sup>. During the chronic phase, this sole genetic abnormality (chromosomal translocation Ph<sup>+</sup>: t(9;22)(q34;q11)) at the stem cell level causes increased proliferation of myeloid cells without loss of their capacity to differentiate. Without treatment, most patients progress to the blast phase when additional oncogenic mutations result in a fatal acute leukaemia made of proliferating immature cells. Imatinib mesylate and other tyrosine kinase inhibitors (TKIs) that target the kinase activity of BCR-ABL have improved patient survival markedly. However, fewer than 10% of patients reach the stage of complete molecular response (CMR), defined as the point when BCR-ABL transcripts become undetectable in blood cells<sup>3</sup>. Failure to reach CMR results from the inability of TKIs to eradicate quiescent CML leukaemia stem cells (LSCs)<sup>2-4</sup>. Here we show that the residual CML LSC pool can be gradually purged by the glitazones, antidiabetic drugs that are agonists of peroxisome proliferator-activated receptor- $\gamma$  (PPAR $\gamma$ ). We found that activation of PPAR $\gamma$  by the glitazones decreases expression of STAT5 and its downstream targets HIF2 $\alpha$ <sup>5</sup> and CITED2<sup>6</sup>, which are key guardians of the quiescence and stemness of CML LSCs. When pioglitazone was given temporarily to three CML patients in chronic residual disease in spite of continuous treatment with imatinib, all of them achieved sustained CMR, up to 4.7 years after withdrawal of pioglitazone. This suggests that clinically relevant cancer eradication may become a generally attainable goal by combination therapy that erodes the cancer stem cell pool.

Cell division tracking with carboxyfluorescein diacetate-succinimide ester (CFSE) indicates that non-cycling CML cells are poorly sensitive to TKIs<sup>7,8</sup> and that the quiescent TKI-resistant subpopulation is enriched in CD34<sup>+</sup>38<sup>-</sup> cells<sup>9</sup>. CML LSCs are hence similar to normal quiescent haematopoietic stem cells (HSCs), although they are cytokine-independent<sup>7</sup>. Because failure to reach CMR occurs even when BCR-ABL remains sensitive to TKIs<sup>2</sup>, we searched for possible 'non-oncogene addiction (NOA)' of CML LSCs as a novel therapeutic target. NOA indicates that a given malignant cell is abnormally sensitive to quantitative variations in an otherwise normal molecular pathway<sup>10</sup>.

We previously reported that the Nef proteins of the immunodeficiency viruses impair haematopoiesis by activating peroxisome

proliferator-activated receptor gamma (PPAR $\gamma$ )<sup>11</sup>. This effect was reproduced by the thiazolidinediones, a class of synthetic PPAR $\gamma$  ligands (Extended Data Fig. 1a), although it is compensated in individuals with otherwise normal haematopoiesis<sup>12</sup>. We then became intrigued with our observation that the CML cell line K562 is particularly sensitive to Nef and thiazolidinediones<sup>11</sup>. The involvement of PPAR $\gamma$  was also more recently reported in haematopoietic stress response<sup>13</sup>.

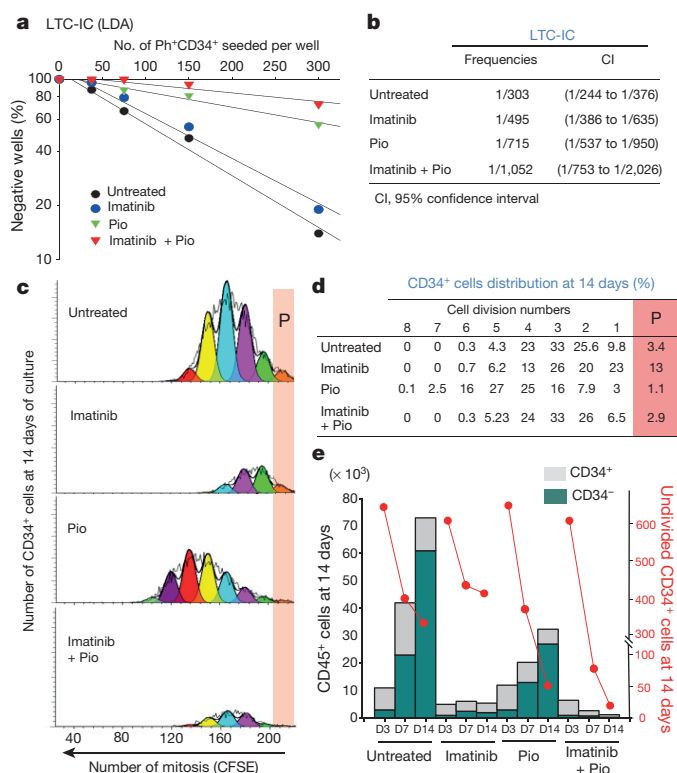
We turned to a cohort of 29 chronic phase (CP) CML patients at diagnosis whose CD34<sup>+</sup> cells were >95% Ph<sup>+</sup>. Combining imatinib and pioglitazone showed evidence of synergy with a decrease in the number of colony-forming cells (CFC) sixfold more pronounced ( $P < 0.0001$ ) than with imatinib alone (Extended Data Fig. 2a). A similar trend was observed when normal CD34<sup>+</sup> cells were transduced with a lentiviral vector expressing p210 BCR-ABL (Extended Data Fig. 2b). Whereas imatinib alone was unable to reduce significantly the frequency of CP-CML long term culture-initiating cells (LTC-ICs) ( $P = 0.067$ ), we found that pioglitazone was able to do so, either as a single agent by 2.4-fold ( $P = 0.008$ ) or with an improved effect by 3.5-fold in the presence of imatinib ( $P < 0.001$ ) (Fig. 1a, b). Similar results were obtained with the second generation TKI dasatinib or with another thiazolidinedione, rosiglitazone (Extended Data Fig. 2c, d).

CFSE assays were then performed with CP-CML CD34<sup>+</sup> cells in the absence of cytokines (Fig. 1c-e and Extended Data Table 1). Untreated control CP-CML CD34<sup>+</sup> cells proliferated and differentiated actively. Imatinib exposure resulted in the elimination of actively dividing cells but also in the accumulation of viable CFSE-bright CD34<sup>+</sup> cells that never divided ('P') or had divided only once (Fig. 1d). Pioglitazone alone was less effective than imatinib to deplete the bulk of dividing CML cells but triggered exit from quiescence (Fig. 1c-e and Extended Data Table 1). Combining pioglitazone with either imatinib or dasatinib acted in synergy to deplete both proliferating and non-proliferating cells (Fig. 1c-e, Extended Data Table 1 and Extended Data Fig. 2e). Imatinib alone was effective at decreasing the number of Ph<sup>+</sup> CD34<sup>+</sup> CD38<sup>+</sup> progenitors but failed to reduce the more immature CD34<sup>+</sup> CD38<sup>-</sup> population, opposite to pioglitazone alone (Extended Data Fig. 3b).

We then investigated the possible molecular pathways that mediate pioglitazone activity against CML LSCs. We previously reported that PPAR $\gamma$  is a negative transcriptional regulator of STAT5 (A and B)<sup>11</sup>. STAT5 is known to be critical for maintenance and fitness of both normal HSCs<sup>14</sup> and CML cells, where STAT5 is activated upon direct phosphorylation by the BCR-ABL kinase<sup>15</sup>. STAT5 expression levels were abnormally high in both total CP-CML CD34<sup>+</sup> cells and quiescent LSC (Fig. 2a). In CFSE-bright cells (that is, P and 1 division of

<sup>1</sup>CEA, Institute of Emerging Diseases and Innovative Therapies (IMETI), F-92265 Fontenay-aux-Roses, France. <sup>2</sup>Département de biologie médicale, Hôpital Mignot, F-78150 Le Chesnay, France. <sup>3</sup>Unité de Biologie Cellulaire, UMR-S-940 Institut Universitaire d'Hématologie, Hôpital Saint Louis, F-75010 Paris, France. <sup>4</sup>Service d'Endocrinologie et de Diabétologie, Hôpital Mignot, F-78150 Le Chesnay, France. <sup>5</sup>CIRI, International Center for Infectiology Research, EVIR team, Inserm, U1111, CNRS, UMR5308, Université de Lyon-1, ENS de Lyon, 69007 Lyon, France. <sup>6</sup>Inserm, U895, Centre de Médecine Moléculaire (C3M), équipe 3, 06204 Nice, France. <sup>7</sup>Laboratoire d'hématologie, Centre Hospitalier de Versailles, F-78150 Le Chesnay, France. <sup>8</sup>Unité de Pharmacologie, Service de Biologie Médicale, Centre Hospitalier de Versailles, F-78150 Le Chesnay, France. <sup>9</sup>Service d'Hématologie et d'Oncologie, Hôpital Mignot, Université Versailles Saint-Quentin-en-Yvelines, F-78150 Le Chesnay, France. <sup>10</sup>Inserm, Institute of Emerging Diseases and Innovative Therapies (IMETI), F-92265 Fontenay-aux-Roses, France. <sup>11</sup>Genetics Division, Brigham & Women's Hospital and Harvard Medical School, Boston, Massachusetts 02115, USA. <sup>12</sup>Hematology Division, Ramathibodi Hospital and Mahidol University, 10400 Bangkok, Thailand.

\*These authors contributed equally to this work.

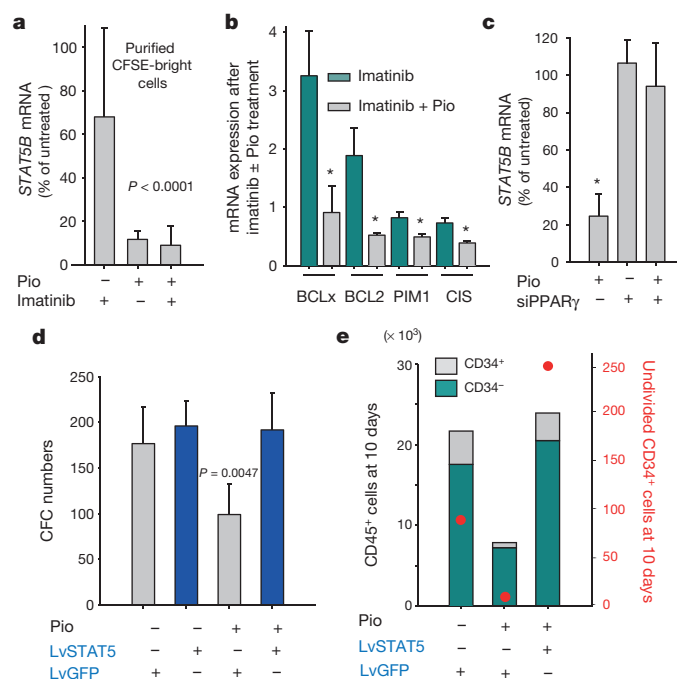


**Figure 1 | Pioglitazone purges quiescent CML stem cells.** **a**, Limited dilution analysis (LDA) of CML LSCs by LTC-IC assay. Pio, pioglitazone. **b**, LTC-IC frequencies calculated from **a** ( $n = 4$ ). **c**, CFSE analysis (patient 4) after liquid culture in serum-free medium without cytokines. P (red), colcemid arrested 'parent-cells'. **d**, Distribution (%) of CD34<sup>+</sup> cells in each mitosis peak shown in **c**. **e**, Identical culture conditions as in **c**, but for patient 2. Left scale (black) and histograms show cell counts. Right scale (red) and red dots and lines show the number of undivided CD34<sup>+</sup> cells (P in CFSE assay). Also see Extended Data Table 1 ( $n = 6$ ). See statistics in Methods.

CP-CML CD34<sup>+</sup> cells) purified at 14 days of culture without cytokines, *STAT5B* messenger RNA levels decreased by 8.5-fold ( $P < 0.0001$ ), 1.5-fold ( $P = 0.08$ ) and 10.5-fold ( $P < 0.0001$ ) in the presence of pioglitazone, imatinib and the drug combination, respectively (Fig. 2a). Similar values were obtained for *STAT5A* (not shown).

We then compared mRNA levels of four known *STAT5* target genes. Addition of pioglitazone to imatinib significantly reduced expression of *BCL2L1* (also known as *BCL-X<sub>L</sub>*)<sup>16</sup> (3.3-fold), *BCL2*<sup>16</sup> (4.8-fold), *PIM1*<sup>17</sup> (1.6-fold) and *CISH* (also known as *CIS*)<sup>18</sup> (1.6-fold), thus suggesting that imatinib alone is not able to inhibit *STAT5* transcriptional activity to completion (Fig. 2b). Supplementary studies with the bromodomain inhibitor JQ1<sup>19</sup> confirmed the pivotal role played by *STAT5* in CML LSCs (Supplementary Data and Extended Data Fig. 2f).

The effect of pioglitazone was negated by a short interfering RNA against *PPAR $\gamma$*  (also known as *PPARG*) mRNA (Fig. 2c and control Extended Data Fig. 1b, c). Decreased clonogenicity of CP-CML CD34<sup>+</sup> cells in the presence of pioglitazone was abolished when *STAT5B* was overexpressed after lentiviral transfer (Fig. 2d). We observed that imatinib acts rapidly (in minutes) by preventing *STAT5* phosphorylation, whereas pioglitazone acts slowly (in days) by decreasing *STAT5* protein levels (Extended Data Fig. 3a, c and d), owing to the known long half-life of *STAT5* protein in spite of the rapid decrease in its mRNA levels<sup>15</sup>. Pioglitazone activity in CFSE assays with CP-CML cells was abrogated when *STAT5B* was overexpressed by lentiviral transfer (Fig. 2e, Extended Data Fig. 4a–e and Extended Data Fig. 1c). Importantly, pioglitazone was found to be more inhibitory/cytotoxic for CML LTC-IC than for normal LTC-IC (Extended Data Fig. 5).

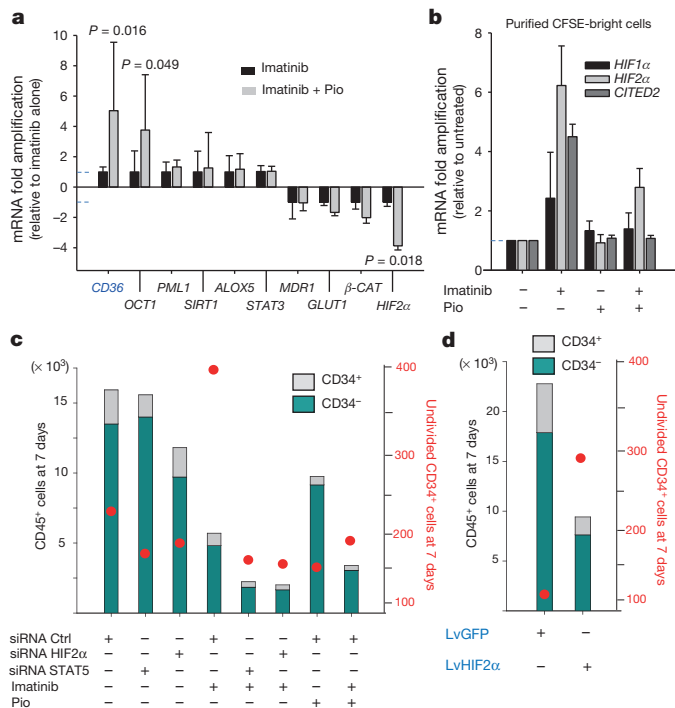


**Figure 2 | Pioglitazone targets the PPAR $\gamma$ -*STAT5* pathway in CML LSCs.** **a**, Normalized *STAT5B* quantitative PCR with reverse transcription (RT-qPCR) on CFSE-bright cells (that is, P and 1 division) at 14 days of culture. **b**, Percent mRNA expression of *STAT5* target genes in CFSE bright cells after drug exposure. **c**, CP-CML CD34<sup>+</sup> cells cultured with an anti-PPAR $\gamma$  siRNA before RT-qPCR. **d**, Colony-forming cell (CFC) assays with CP-CML CD34<sup>+</sup> cells after transduction with enhanced green fluorescent protein (eGFP; negative control) or a *STAT5B*-expressing lentivectors (Lv). **e**, Absolute cell count together with CFSE analysis (patient 2 in triplicate). Red dots, undivided CD34<sup>+</sup> cells (P in CFSE assay). Data show means  $\pm$  s.d.,  $n = 5$ . See statistics in Methods.

We then examined, in 7-day cultures without cytokines of CP-CML CD34<sup>+</sup> cells from 11 patients, mRNA expression levels for 9 putative downstream transcriptional targets of *STAT5* and/or *PPAR $\gamma$*  (Fig. 3a). These included *OCT1* (also known as *POU2F1*)<sup>20</sup>, *PML*, *SIRT1*, *ALOX5*, *STAT3*, *MDR1* (also known as *ABCB1*), *GLUT1* (also known as *SLC2A1*),  $\beta$ -catenin (also known as *CTNNB1*) and *HIF2 $\alpha$*  (also known as *EPAS1*)<sup>17</sup>. *CD36*, known to be upregulated by *PPAR $\gamma$*  agonists, was used as a positive control<sup>21</sup>. Only *OCT1* and *HIF2 $\alpha$*  expression levels were significantly altered after culture in the presence of pioglitazone + imatinib versus imatinib alone (Fig. 3a). Although upregulation of *OCT1* expression may increase the cellular uptake of imatinib<sup>20</sup>, we found that erosion of the CP-CML LSC pool was not improved in the presence of imatinib alone when *OCT1* was overexpressed by lentiviral transfer (Extended Data Fig. 6).

In contrast to *OCT1*, *HIF2 $\alpha$*  was downregulated by pioglitazone (Fig. 3a). Importantly, *HIF2 $\alpha$*  and to a lesser degree *HIF1 $\alpha$*  were found upregulated in imatinib-resistant CFSE-bright cells (P and 1 cell division), while pioglitazone counteracted this phenomenon (Fig. 3b). *CITED2*, a key gene of HSC 'stemness'<sup>6</sup>, known to be upregulated by *HIF2 $\alpha$*  followed the same trend (Fig. 3b). The viability of undivided and imatinib-resistant CP-CML CD34<sup>+</sup> cells required *HIF2 $\alpha$*  expression (Fig. 3c, Extended Data Fig. 7a–e and Extended Data Fig. 1c), and forced expression of *HIF2 $\alpha$*  in normal human cord blood CD34<sup>+</sup> cells increased the compartment of quiescent cells (Fig. 3d, Extended Data Fig. 7f–i). Furthermore, forced expression of BCR-ABL in normal human CD34<sup>+</sup> cells induced *HIF2 $\alpha$*  in a *STAT5*-dependent manner (Extended Data Fig. 8a). Similarly, forced expression of constitutively active mutants of *STAT5A* or *STAT5B* also induced *HIF2 $\alpha$*  expression (Extended Data Fig. 8b). In both cases, induction of *HIF2 $\alpha$*  was associated with upregulated expression of its target gene, *CITED2* (Extended



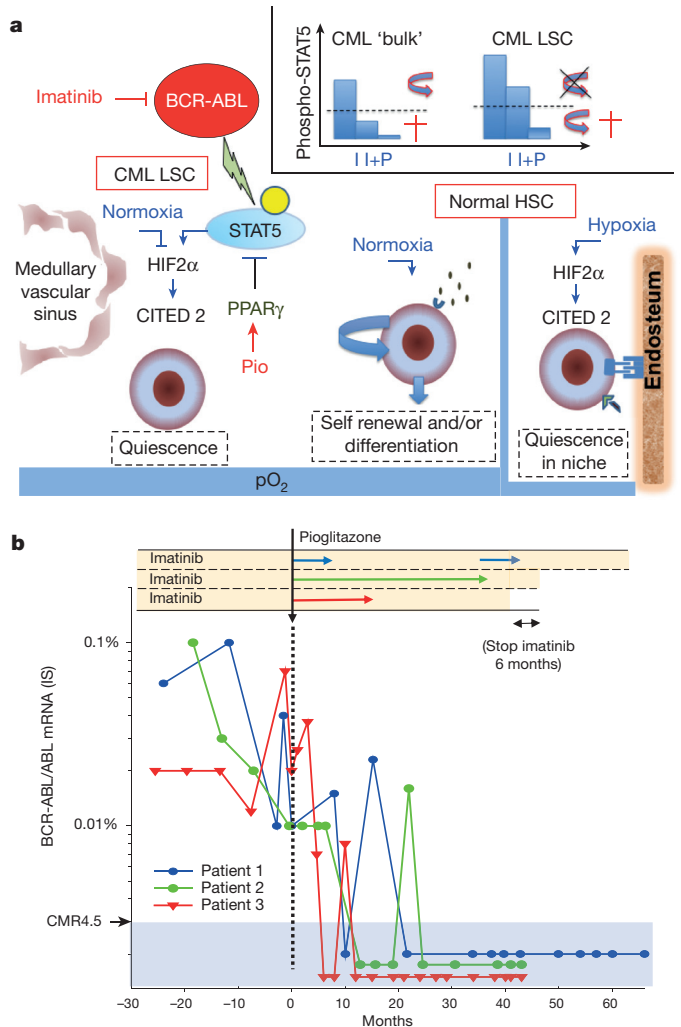


**Figure 3 | Expression of target genes in CP-CML cells exposed to pioglitazone and imatinib.** **a**, RT-qPCR assays on CP-CML CD34<sup>+</sup> cells after 7 days without serum or cytokines. Data show means  $\pm$  s.d.,  $n = 11$ . **b**, RT-qPCR assays in purified 12–14 days CFSE-bright cells (that is, P and 1 division). Data show means  $\pm$  s.d.,  $n = 6$ . **c**, Absolute cell count together with CFSE analysis of representative CP-CML patient 8 (triplicate). Undivided cells (red dots). Also see Extended Data Fig. 7 and Extended Data Fig. 1c. **d**, Same as in **c**, but after transduction of cord blood CD34<sup>+</sup> cells with lentivectors (Lv) (triplicate). Undivided cells (red dots). See statistics in Methods.

Data Fig. 8a, b). Because *CITED2* is a known master gene of HSC quiescence that regulates stemness-associated genes such as *BMI1*<sup>22</sup>, *HES1*<sup>23</sup> and *p57* (also known as *CDKN1C*)<sup>23</sup>, we studied the expression of these genes in CD34<sup>+</sup> cells from CP-CML patients and in murine Ba/F3 cell lines we generated to express, by means of retroviral vector transduction and ubiquitous promoter driven expression, the constitutively active forms of murine Stat5a or Stat5b I\*6 (H299R, S711F). After 10 days of culture in the presence of imatinib, TKI-resistant CD34<sup>+</sup> cells from CP-CML patients showed an increase in endogenous expression of both *CITED2* itself (4.5-fold) and the known *CITED2* target genes *BMI1* (2.8-fold), *HES1* (3.1-fold) and *p57* (16.5-fold). Addition of pioglitazone fully counteracted said increase in *CITED2*, *BMI1* and *HES1* expression and reduced the increase in *p57* expression by fourfold (Extended Data Fig. 9a). Ba/F3 cell studies corroborated this evidence (Extended Data Fig. 9b and c). Taken together, we propose here that the CML-LSC is critically dependent (NOA) on a PPAR $\gamma$ -STAT5-HIF2 $\alpha$ -CITED2 pathway, directly and effectively inhibited by pioglitazone (Fig. 4a), thus extending the contention that equivalent murine leukaemias are addicted (NOA) to STAT5 (ref. 15).

Because mouse models are poorly suited to investigate CML LSCs<sup>24</sup> and pioglitazone is an approved drug for the treatment of diabetes mellitus type 2 in humans, we initially sought to validate pioglitazone directly on two patients diagnosed with both diabetes and CML who never reached CMR in spite of long-term imatinib treatment. Before filing a formal clinical trial application, we prescribed pioglitazone off-label and under approved informed consent to a third CML patient, this time non-diabetic, who never reached CMR either under long-term imatinib therapy (Fig. 4b).

Pioglitazone was added to the treatment after 5, 6 and 4 years of uninterrupted imatinib therapy for patients 1, 2 and 3, respectively. None of the 3 patients ever reached CMR before introduction of



**Figure 4 | Pioglitazone induces complete and sustained molecular response (CMR) in CML patients.** **a**, Model of CML LSC addiction to the PPAR $\gamma$ -STAT5-HIF2 $\alpha$  pathway. (Top insert) for the bulk of dividing CML cells, imatinib (I) alone is able to bring phospho-STAT5 levels below a threshold (dotted line) at which apoptosis occurs (cross). For CML LSCs, only the combination of imatinib and pioglitazone (I+P) is able to bring cells below a threshold at which cells leave their state of quiescence before undergoing apoptosis. **b**, RT-qPCR assays for *BCR-ABL/ABL* on nucleated blood cells from the first three patients. See Supplementary Information for details.

pioglitazone. Pioglitazone was added to the treatment of patient 1 during two brief exposures of 10 and 8 months each with an interval of 28 months (Fig. 4b). CMR was achieved 10 months after initial pioglitazone addition, and patient 1 has remained in CMR for at least 56 months, the last time-point collected for this study, which is 53 months (4.5 years) after first stopping pioglitazone administration (Fig. 4b). For patient 2, CMR was obtained after 1 year of pioglitazone addition and maintained for 32 months at which time they withdrew (Fig. 4b). For patient 3, CMR was achieved after 6 months of pioglitazone addition. At this time point, the level of *STAT5* mRNA in CD34<sup>+</sup> cells from the bone marrow of patient 3 was decreased by 11.9 fold. Patient 3 has remained in CMR for at least 38 months, the last time-point collected for this study, which is 28 months after stopping pioglitazone administration (Fig. 4b). Furthermore, patient 3 decided to stop imatinib for the last 6 months of the aforementioned observation period and has remained in CMR during this period without any treatment (Fig. 4b and Supplementary Data).

Regulatory approval was then obtained for multi-centre Phase II clinical trials, and the first (EudraCT 2009-011675-79) aimed at assessing the short-term cumulative incidence of CMR conversion

for patients who never reached CMR under imatinib alone (<https://www.clinicaltrialsregister.eu/ctr-search/trial/2009-011675-79/FR#E>). Scoring by quantitative PCR was performed over the course of the first 12 months after trial initiation during concurrent and brief exposure (3 to 12 months) to the imatinib-pioglitazone combination<sup>25</sup>. Out of 24 assessable patients, the cumulative incidence rate in the treated group reached 57% versus 27% ( $P = 0.02$ ) for an historical group of patients having received imatinib alone, thus indicating that clinical evidence of efficacy can already be detected even after very brief treatment and early analysis. Post-trial follow-up confirmed stability of CMR status in data collected to date. Therapy with pioglitazone was accompanied by a stable reduction of STAT5 mRNA in patient samples as early as month 6 (2.3-fold,  $P = 0.0003$ ) and by a reduction of the clonogenic potential of bone marrow CD34<sup>+</sup> cells (1.54-fold,  $P = 0.0003$ ).

Although both imatinib and pioglitazone decrease STAT5 activity, they act by different mechanisms. Imatinib inhibits STAT5 activation by BCR-ABL phosphorylation, whereas pioglitazone decreases STAT5 expression. It seems that imatinib alone is sufficient to induce effective clearance of the bulk of more differentiated CML cells, but fails to bring STAT5 activity below a threshold for CML LSC to exit from quiescence and to undergo subsequent apoptosis. Pioglitazone is effective at doing so in synergy with imatinib (Fig. 4a, top insert). As in this example with CML, progressive erosion of cancer stem cell pools may prove ultimately achievable pharmacologically, bringing hope of obtaining cancer eradication in a variety of human malignancies by combination therapy.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 17 March 2014; accepted 28 July 2015.**

**Published online 2 September 2015.**

- Nguyen, L. V., Vanner, R., Dirks, P. & Eaves, C. J. Cancer stem cells: an evolving concept. *Nature Rev. Cancer* **12**, 133–143 (2012).
- Chomel, J. C. & Turhan, A. G. Chronic myeloid leukemia stem cells in the era of targeted therapies: resistance, persistence and long-term dormancy. *Oncotarget* **2**, 713–727 (2011).
- de Lavallade, H. *et al.* Imatinib for newly diagnosed patients with chronic myeloid leukemia: incidence of sustained responses in an intention-to-treat analysis. *J. Clin. Oncol.* **26**, 3358–3363 (2008).
- Corbin, A. S. *et al.* Human chronic myeloid leukemia stem cells are insensitive to imatinib despite inhibition of BCR-ABL activity. *J. Clin. Invest.* **121**, 396–409 (2011).
- Hu, C. J., Sataur, A., Wang, L., Chen, H. & Simon, M. C. The N-terminal transactivation domain confers target gene specificity of hypoxia-inducible factors HIF-1 $\alpha$  and HIF-2 $\alpha$ . *Mol. Biol. Cell* **18**, 4528–4542 (2007).
- Du, J. & Yang, Y. C. Cited2 in hematopoietic stem cell function. *Curr. Opin. Hematol.* **20**, 301–307 (2013).
- Graham, S. M. *et al.* Primitive, quiescent, Philadelphia-positive stem cells from patients with chronic myeloid leukemia are insensitive to STI571 *in vitro*. *Blood* **99**, 319–325 (2002).
- Jørgensen, H. G., Allan, E. K., Jordanides, N. E., Mountford, J. C. & Holyoake, T. L. Nilotinib exerts equipotent antiproliferative effects to imatinib and does not induce apoptosis in CD34<sup>+</sup> CML cells. *Blood* **109**, 4016–4019 (2007).
- Copland, M. *et al.* Dasatinib (BMS-354825) targets an earlier progenitor population than imatinib in primary CML but does not eliminate the quiescent fraction. *Blood* **107**, 4532–4539 (2006).
- Luo, J., Solimini, N. L. & Elledge, S. J. Principles of cancer therapy: oncogene and non-oncogene addiction. *Cell* **136**, 823–837 (2009).
- Prost, S. *et al.* Human and simian immunodeficiency viruses deregulate early hematopoiesis through a Nef/PPAR $\gamma$ /STAT5 signaling pathway in macaques. *J. Clin. Invest.* **118**, 1765–1775 (2008).
- Berria, R. *et al.* Reduction in hematocrit and hemoglobin following pioglitazone treatment is not hemodilutional in Type II diabetes mellitus. *Clin. Pharmacol. Ther.* **82**, 275–281 (2007).
- Avagyan, S., Aguilo, F., Kamezaki, K. & Snoeck, H. W. Quantitative trait mapping reveals a regulatory axis involving peroxisome proliferator-activated receptors, PRDM16, transforming growth factor- $\beta$ 2 and FLT3 in hematopoiesis. *Blood* **118**, 6078–6086 (2011).
- Wang, Z., Li, G., Tse, W. & Bunting, K. D. Conditional deletion of STAT5 in adult mouse hematopoietic stem cells causes loss of quiescence and permits efficient nonablative stem cell replacement. *Blood* **113**, 4856–4865 (2009).
- Hoelbl, A. *et al.* Stat5 is indispensable for the maintenance of bcr/abl-positive leukaemia. *EMBO Mol. Med.* **2**, 98–110 (2010).
- Kieslinger, M. *et al.* Antiapoptotic activity of Stat5 required during terminal stages of myeloid differentiation. *Genes Dev.* **14**, 232–244 (2000).
- Fatrai, S., Wierenga, A. T., Daenen, S. M., Vellenga, E. & Schuringa, J. J. Identification of HIF2 $\alpha$  as an important STAT5 target gene in human hematopoietic stem cells. *Blood* **117**, 3320–3330 (2011).
- Matsumoto, A. *et al.* CIS, a cytokine inducible SH2 protein, is a target of the JAK-STAT5 pathway and modulates STAT5 activation. *Blood* **89**, 3148–3154 (1997).
- Liu, S. *et al.* Targeting STAT5 in hematologic malignancies through inhibition of the bromodomain and extra-terminal (BET) bromodomain protein BRD2. *Mol. Cancer Ther.* **13**, 1194–1205 (2014).
- Wang, L., Giannoudis, A., Austin, G. & Clark, R. E. Peroxisome proliferator-activated receptor activation increases imatinib uptake and killing of chronic myeloid leukemia cells. *Exp. Hematol.* **40**, 811–819 (2012).
- Szanto, A. & Nagy, L. Retinoids potentiate peroxisome proliferator-activated receptor gamma action in differentiation, gene expression, and lipid metabolic processes in developing myeloid cells. *Mol. Pharmacol.* **67**, 1935–1943 (2005).
- Chen, Y., Haviernik, P., Bunting, K. D. & Yang, Y. C. Cited2 is required for normal hematopoiesis in the murine fetal liver. *Blood* **110**, 2889–2898 (2007).
- Du, J. *et al.* HIF-1 $\alpha$  deletion partially rescues defects of hematopoietic stem cell quiescence caused by Cited2 deficiency. *Blood* **119**, 2789–2798 (2012).
- Koschmieder, S. & Schemionek, M. Mouse models as tools to understand and study BCR-ABL1 diseases. *Am. J. Blood Res.* **1**, 65–75 (2011).
- Rousselot, P. *et al.* Targeting STAT5 expression resulted in molecular response improvement in patients with chronic phase CML treated with imatinib. *ASH Annual Meeting Abstracts*, (2012).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank C. Costa, V. Tran Chau, F. Goullieux, A. Krief, P. Raynal, C. Terré, S. Tabore and T. Andrieu for their experimental contributions. This work was supported by the Association Laurette Fugain, Paris, France, by the Association pour la Recherche sur le Cancer, Villejuif, France to S.P., P.R. and P.L. and by the Chaire industrielle de l'Agence Nationale pour la Recherche (ANR) to P.L.

**Author Contributions** S.P. lead the project, designed and performed experiments, and analysed data. P.L. and S.Ch. designed experiments and analysed data. P.L. wrote the paper. F.R., M.S., Y.O., J.S., E.V., V.R., B.M. and G.M. contributed experimentally. P.R., J.-P.B., C.C. and S.Ca. contributed clinically. P.R., S.Ch. and P.L. have contributed equally to this work; P.R. in a clinical capacity, S.C. and P.L. in a scientific capacity.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.P. ([stephane.prost@cea.fr](mailto:stephane.prost@cea.fr)) and P.L. ([pleboulch@rics.bwh.harvard.edu](mailto:pleboulch@rics.bwh.harvard.edu)).

## METHODS

**Reagents.** For *in vitro* assays, PPAR $\gamma$  agonists were provided by Cayman Chemical (PPAR $\gamma$ -PAK; Bertin-pharma). Imatinib mesylate was provided by Novartis and was used at 1  $\mu$ M in culture, a well-established inhibitory concentration *in vitro* that also approaches the achievable drug level in patients' plasma. Dasatinib and JQ1 were provided by Bristol-Myers Squibb and Sigma and were used in culture at 0.146  $\mu$ M and 1  $\mu$ M, respectively. Murine pro-B cell line Ba/F3 and human chronic myelogenous leukaemia cell line K562 were provided by the American Type Culture Collection (ATCC; Ref. CRL-12015 and CCL-243, respectively). These cell lines were tested for mycoplasma contamination every 3 months using Venor Gem Advance Pre-aliquoted Mycoplasma Detection Kit (Minerva biolabs).

**Cell culture and proliferation assays.** CD34 $^{+}$  cells from patients in CP-CML at diagnosis or umbilical cord blood were immunoselected (CD34 microBead Kit, Miltenyi Biotec) according to the manufacturer's instructions. Enrichment for CD34 $^{+}$  cells was ascertained by flow cytometry using an anti-CD34 monoclonal antibody (clone 581; BD Pharmingen). Ph1 $^{+}$ -CD34 $^{+}$  cells were cultured in serum free medium (SFM) StemSpan (StemCell Technologies) without growth factors.

**Colony forming cell (CFC) and long term culture-initiating cell (LTC-IC) assays.** For CFC assays, CD34 $^{+}$  cells were suspended ( $1 \times 10^4$ ) in 3 ml of alpha-MEM based methylcellulose medium (GF H4434, StemCell Technologies). Cells were scored and collected after 14 days incubation at 37 °C and 5% CO $_2$ . After scoring, colonies were washed with PBS and kept frozen in RNAlater (Invitrogen) for subsequent analysis. LTC-IC with limiting dilution assays (LDA) were performed in StemSpan SFM (Stemcell technologies) on irradiated MS5 monolayers at several dilutions of CD34 $^{+}$  cells (300, 150, 75, or 37 cells per well for Ph1 $^{+}$  CD34 $^{+}$  cells and 200, 100, 50, or 25 cells per well for CD34 $^{+}$  from healthy donors) in 96-well plates with 16 replicate wells per concentration. After five weeks with weekly change of one half medium volume, all cells were transferred in alpha-MEM based methylcellulose medium (GF H4434, Stemcell technologies) to determine the total clonogenic cell content of each LTC. LTC-IC frequencies were determined using the L-Cal software (Stemcell technologies).

**Flow cytometry.** The following antibodies were used: fluorescein isothiocyanate (FITC)-conjugated IgG1 (clone 679.1Mc7, Beckman Coulter), Alexa Fluor 488-conjugated-IgG1 (clone MOPC-21, BD Pharmingen), allophycocyanin (APC)-IgG1 (clone MOPC-21, BD Pharmingen), peridinin chlorophyll protein-cyanin 5.5 (PerCP-Cy5.5)-conjugated IgG1 (clone X40, BD Pharmingen), phycoerythrin cyanin (PE-Cy7)-conjugated IgG1 (clone MOPC-21, BD Pharmingen), (PerCP-Cy5.5)-conjugate CD45 (clone 2D1, BD Pharmingen), (APC)-conjugated CD34 (clone 581, BD Pharmingen), (PE-Cy7)-conjugated CD38 (clone HB7, BD Pharmingen), Alexa Fluor 488-conjugated anti-STAT5 (pY694) (clone 47, BD Pharmingen), (PE)-conjugated anti-GLUT1 (FAB1418P, R&D systems). For all experiments, cell viability was assessed using SYTOX Blue dead cell stain (Invitrogen Life Technologies).

**Intracellular STAT5 phosphorylation assays.** In brief,  $3 \times 10^5$  K562 cells per ml cultured in complete Dulbecco's modified Eagle medium supplemented with 10% fetal calf serum (PAA) alone and with or without pioglitazone (10  $\mu$ M) or imatinib (1  $\mu$ M) at 37 °C in 5% CO $_2$  were harvested at variable time as indicated. Cells were fixed and permeabilized using Cytotfix/Cytoperm kit (BD Pharmingen) and stained with Alexa Fluor 488-anti-phospho-STAT5 monoclonal antibody (BD Phosflow) or Alexa Fluor 488-isotype-matched control to obtain fluorescence minus comparative in each experiment. Analysis was carried on a minimal number of 50,000 events in the viable cell gate. The delta mean fluorescence intensity of p-STAT5 after drug treatment (p-STAT5MFI) was determined as follow: (untreated cells p-STAT5 MFI – non-treated cells isotype-control MFI) – (drug treated cells p-STAT5 MFI – drug treated cells isotype-control MFI).

**CSFE assays.** Fresh CD34 $^{+}$ -enriched cells were stained with 2  $\mu$ M of 5- (and 6-) carboxyfluorescein diacetate succinimidyl diester (CFSE, Invitrogen). Cells were then cultured (seeded  $5.10^5$  per ml) in SFM StemSpan (StemCell Technologies) without growth factors and with or without pioglitazone (10  $\mu$ M) or imatinib (1  $\mu$ M). Cells cultured in the presence of Colcemid (100 ng ml $^{-1}$ , Invitrogen Life Technologies) were used to establish the range of fluorescence exhibited by cells that had not divided during post-labelling incubation. Cells were harvested at variable time points as indicated, collected in BD Truocount tubes for absolute count (BD Biosciences) and labelled with anti-CD45 and anti-CD34. Then, cells were diluted in 1 ml of phosphate-buffered saline (PBS, Invitrogen Life Technologies) containing 2% fetal calf serum (PAA) and stained for viability. All analyses were carried out on a BD FACS Canto2 Flow Cytometer.

**DNA synthesis assay.** Cell proliferation rate was measured by incorporation of 5-ethynyl-2'-deoxyuridine (EdU), a thymidine nucleoside analogue, in DNA during active DNA synthesis (two hours). Staining was performed according to the manufacturer's protocol (Click-iT EdU Flow Cytometry Assay Kit, Invitrogen). All analyses were carried out on a BD FACS Canto2 Flow Cytometer.

**RNA extraction and RT-qPCR analysis.** RNA was extracted from  $2 \times 10^5$  cells using RNeasy-4PCR (Ambion). Reverse transcription was carried out for 1 h at 42 °C using SuperScript Vilo cDNA Synthesis kit (Invitrogen Life Technologies) according to the manufacturer's instructions. Real-time PCR was performed in an iCycler thermocycler (CFX, Bio-Rad). The primers and probes sequences for *GAPDH*, *STAT5B*, *STAT5A*, *BCR-ABL*, *ABL*, *HIF1 $\alpha$* , *CITED2*, *OCT1*<sup>20</sup>, *MDR1*<sup>26</sup>, *SIRT1*<sup>27</sup>, *STAT3*<sup>28</sup>, *ALOX5*<sup>29</sup>, *GLUT1*<sup>30</sup>,  $\beta$ -catenin<sup>31</sup>, *PML*<sup>32</sup>, *HIF2 $\alpha$* <sup>17</sup>, *Bcl-X $_L$* , *Bcl-2*, *PIM-1*, *CIS*<sup>19</sup>, *BMI1*<sup>22</sup>, *HES-1*<sup>23</sup>, *p57 (CDKN1C)*<sup>23</sup>, *CD36*<sup>21</sup> (known to be upregulated by PPAR $\gamma$  agonists, was used as a positive control) are reported in Supplementary Table 1. The primer pairs used with TaqMan Gene Expression Master mix (Applied Biosystems) and iQ Supermix SYBR GRN (Bio-Rad) are listed in Supplementary Table 1a and Supplementary Table 1b, respectively. The comparative  $C_T$  method ( $\Delta\Delta C_T$ ) was used to compare gene expression levels between the different culture conditions (relative to *GAPDH*).

**BCR-ABL/ABL quantification.** qPCR experiments were performed on cDNA using the 7000 Sequence Detection System (Applied Biosystems). The *BCR-ABL/ABL* ratio was determined using FusionQuant standards (Ipsogen) according to the Europe Against Cancer protocol<sup>33</sup>. CMR is defined as undetectable minimal residual disease (negative *BCR-ABL* transcripts) while showing a sensitivity level of at least 40,000 amplified copies of the *ABL* control gene, that is to say more than 4.5 log reduction by standardized International Scale (IS) RT-qPCR (that is, *BCR-ABL/ABL*<sup>IS</sup> mRNA ratio < 0.0025%); relapse from CMR is to be declared when at least 2 consecutive positives occur 6 months apart. These criteria are consistent with the level of sensitivity routinely applied within laboratories participating in the French GBMHM Network (Group of Molecular Biologists for Hematological Malignancies).

**Interphase FISH probe assay.** Fluorescent *in situ* hybridization (FISH) was performed on interphase nuclei, following standard procedures and using specific probe for the t(9;22) (MetaSystems, Germany). The probe is designed as a dual-fusion assay. The red labelled probe detects an extended region at the *ABL1* locus on 9q34 and a green labelled probe hybridizes specifically to regions at the *BCR* gene on 22q11. Preparations were counterstained with 4,6-diamidino-phenylindole (DAPI) and a minimum of 50 interphase nuclei were examined. Results were recorded using a fluorescence microscope (Nikon) fitted with appropriate filters, and digital-imaging software Lucky (CaryoSystems, France).

**Western blot analysis.** For STAT5 protein analysis, K562 cells ( $2.5 \times 10^5$ ) were lysed in RIPA lysis buffer on ice. Whole-cell extracts were boiled for 5 min in Laemmli sample buffer and subjected to SDS-PAGE in 4–12% acrylamide gels (Nupage, Invitrogen Life Technologies). Proteins were transferred to Hybond N+ filters (Amersham). Membranes were probed with the following antibodies: STAT5 (sc-1656), ACTIN (sc-8432), PPAR $\gamma$  (H-100;sc-7196), goat anti-mouse IgG-HRP (sc-2005) (Santa Cruz Biotechnology Inc.) and Anti-HIF2 $\alpha$  (SMC-185C/D). Antibody binding was detected by the enhanced chemiluminescence ECL+ (Amersham).

**Lentiviral vector production and transduction.** *STAT5B* lentiviral vector. The cDNA encoding STAT5B was cloned, sequenced (GenBank accession number DQ267926), and inserted into the SIN-cPPT-PGK-WHV lentiviral transfer vector as previously described<sup>34</sup>. A SIN-cPPT-PGK-eGFP-WHV lentiviral vector was used for control.

**OCT-1 and HIF2 $\alpha$  lentiviral vectors.** OCT-1 (SLC22A1, accession number BC126364) and HIF2 $\alpha$  (EPAS1, accession number BC051338) lentiviral vectors were provided by Applied Biological Materials, Inc. (catalogue nos LV309003 and LV149063).

**Constitutively active murine Stat5a and Stat5b.** Stat5a(1\*6)(H299R, S711F) and Stat5b(1\*6)(H299R, S711F) retroviral vectors were provided by Cell Biolabs, Inc. (catalogue nos RTV-333 and RTV-335).

**shRNA lentiviral vector anti-PPAR- $\gamma$ .** The PPAR- $\gamma$  mRNA pairing sequence 5'-TGTTCCGTGACAATCTGTC-3' (GenBank accession number L40904) was designed and synthesized as follows within an shRNA structure comprising unique restriction sites at each end: sense 5'-GATCTCCTGTTCCGTGACAATCTGTCCTCAAGAGAACAGATTGTACACGGAACATTTTGGAGAAGATTCC-3'; antisense 5'-CTGAGGAATCTTCCAAAAATGTTCCGTGACAATCTGTAAGTTCTCTACAGATTGTACACGGAACAGCA-3'. Oligonucleotides were annealed and ligated into BglIII and XhoI sites of linearized pSuper plasmid. PolIII H1 promoter-shRNA PPAR- $\gamma$  was then subcloned in the pTRIP lentiviral vector. Vectors were produced as previously described<sup>34</sup>.

**BCR-ABL lentiviral vector.** Total RNA from K562 cells was extracted using TRIzol (Invitrogen Life Technologies). Reverse transcription was carried out for 1 h at 50 °C using SuperScript III (Invitrogen Life Technologies). Two independent PCR were performed using *BCR-ABL* F1, 5'-ATGGTGGACCCGGTGGGCTT-3' with *BCR-ABL* R 2831, 5'-CTGCTACCTCTGCACTATGTCACTG-3' and *BCR-ABL* F 2685, 5'-TCCGCTGACCATCAATAAGGA-3' with *BCR-ABL* R 6097, 5'-CTGCTACCTCTGCACTATGTCACTG-3' respectively. Specific amplification



bands were pooled, heated to 95 °C during 3 min and ramp cooled to 25 °C over a period of 45 min. Annealing product was submitted to a third PCR with *LA Taq* DNA polymerase (Takara) using the following primer pair: *BCR-ABL* F1 asc1: 5'-AGGCGCGCCATGGTGGACCCGGTGGGCTT-3' and *BCR-ABL* R 6097 sbf1: 5'-CCTGCAGGCTGCTACCTCTGCACTATGTCACCTG-3'. Amplification product was subcloned into a pCR-XL-TOPO plasmid (Invitrogen Life Technologies) before being inserted into the SIV GAE-SSFV lentiviral transfer vector, followed by DNA sequencing. An SIV GAE-SSFV-eGFP vector was used as a control. The SIV vectors were produced as previously described<sup>34</sup>.

**CD34<sup>+</sup> cell transduction.** Cells were suspended ( $1 \times 10^6$  ml<sup>-1</sup>) in StemSpan (StemCell Technologies, France) supplemented with protamine sulphate (4 µg ml<sup>-1</sup>), SCF (100 ng ml<sup>-1</sup>), FLT-3-L (100 ng ml<sup>-1</sup>), IL-3 (20 ng ml<sup>-1</sup>), and IL-6 (20 ng ml<sup>-1</sup>), in a 96-well plate coated with RetroNectin (Takara Shuzo Co., Japan). Cell suspensions were incubated for 16 h. Lentiviral vectors were then added and cell suspensions incubated for 12 h. Cells were washed twice before being seeded.

**siRNA assays.** siRNA targeting the human *PPAR $\gamma$*  sequence 5'-TGTTCCGTGACAATCTGTC-3' were synthesized (Sigma-Aldrich Proligo). siRNAs targeting the human *STAT5* sequence 5'-AAACTCAGGACCACTTGC-3', human *HIF2 $\alpha$*  sequence 5'-ATTAGAGCAAAGAGTCAGC-3' and murine *Cited2* sequence 5'-CGAGGAAGTGCTTATGTCCTT-3' were synthesized (eurofins MWG/operon). CD34<sup>+</sup> BM cells were transfected with specific siRNA (25 nM) or control siRNA in the presence of Lipofectamine 2000 (Invitrogen) and maintained for 48 h before CFC assay. Control siRNA was purchased from Invitrogen Life Technologies (BLOCK-iT). Transfection efficiency was assessed using a fluorescein-labelled, double-strand RNA duplex (BLOCK-iT FluorescentOligo; Invitrogen).

**Human patients.** Fresh bone marrow from patients with chronic-phase CML at diagnosis, umbilical cord blood cells from healthy donors, and blood or bone marrow samples from diabetes patients and patient given pioglitazone off-label were obtained with informed consent approved by the hospital's Institutional Review Board (Comité de protection des personnes Ile-de-France XI) under approved protocol EudraCT number: 2009-011675-79. Imatinib levels were measured in patients' plasma using a previously described high-performance liquid chromatography (HPLC) method<sup>35</sup>.

**Statistical analysis.** No statistical methods were used to predetermine sample size, the experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment. For culture assays and quantitative real-time PCR, values were calculated as mean  $\pm$  standard deviation for at least three separate experiments performed in triplicate. Paired and unpaired comparisons were made, using the nonparametric Wilcoxon rank test and the Mann-Whitney test, respectively. Limiting dilution analysis was carried out with L-Cal software (StemCell Technologies). All statistical analyses were carried out with StatView software (SAS Institute Inc., Cary, NC).

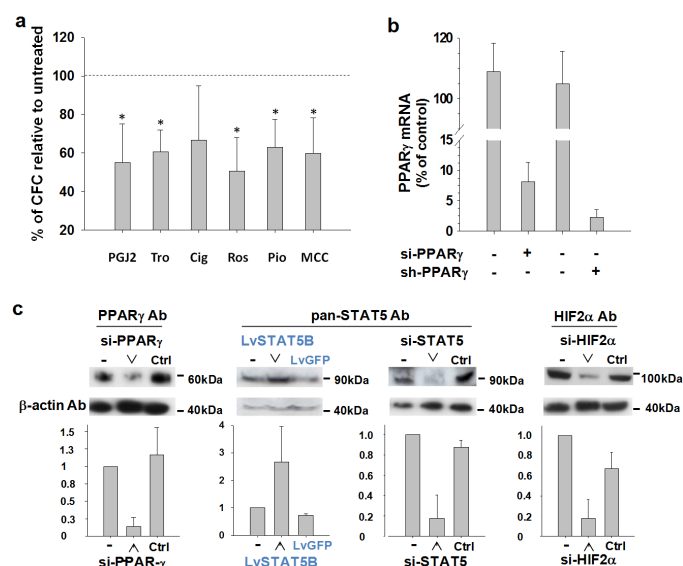
**Statistical information on samples described in figures.** For all culture assays, paired and unpaired comparisons were made using the nonparametric Wilcoxon rank test and the Mann-Whitney test, respectively. Fig. 1a, b, plotted are means for CD34<sup>+</sup> cells from 4 CP-CML patients, 16 replica for each. Data with imatinib alone are not statistically different from those for the untreated control ( $P = 0.067$ ), whereas pioglitazone as a single agent reduced LTC-IC frequencies by 2.4-fold ( $P = 0.008$ ) or by 3.5-fold in combination with imatinib ( $P < 0.001$ ). LTC-IC frequencies were established using L-Cal software (StemCell Technologies). Fig. 1c–e, All patients and statistical analysis are presented in Extended Data Table 1 ( $n = 6$ ). Fig. 2a, *STAT5B* RT-qPCR normalized to *GAPDH* mRNA. Shown are means with standard deviations (s.d.) for 5 CP-CML patients. *STAT5B* mRNA levels decreased by 8.5-fold ( $P < 0.0001$ ), 1.5-fold ( $P = 0.08$ ) and 10.5-fold ( $P < 0.0001$ ) in the presence of pioglitazone, imatinib and the drug combination, respectively. Fig. 2b, Compared to imatinib alone, RT-qPCR analysis shows that addition of pioglitazone induces a significant reduction in mRNA levels by 3.3 and 4.8 fold for *BCL-x<sub>L</sub>* and *BCL-2*, respectively, and 1.6 fold for *PIM1* and *CIS*. mRNA

quantification are normalized to *GAPDH* levels ( $n = 5$  CP-CML patients,  $*P < 0.05$ ). Fig. 2c, Means with s.d. for 5 CP-CML patients. The effect of pioglitazone was negated by an siRNA against *PPAR $\gamma$*  ( $P = 0.043$ ); siRNA validated in Extended Data Fig. 1c. Fig. 2d, Means with s.d. for 5 CP-CML patients. *STAT5* overexpression counteract pioglitazone effect ( $P = 0.0047$ ); *LvSTAT5* validated in Extended Data Fig. 1c. Fig. 3a, results are normalized to *GAPDH* mRNA levels and represented relative to mRNA expression for the "Imatinib alone" condition (means of 11 patients with s.d. for each mRNA assessed). Fig. 3b, results are normalized to *GAPDH* mRNA levels and represented relative to mRNA expression of "untreated cells" (means of 6 patients with s.d.). As compared to untreated controls, cells treated with imatinib alone show a 6.2-fold increase of *HIF2 $\alpha$*  relative to control ( $P < 0.011$ ) and a 4.5-fold increase of *CITED2* ( $P = 0.0277$ ). The addition of pioglitazone reduced the imatinib-mediated *HIF2 $\alpha$*  increase to 2.8-fold ( $P = 0.027$ ). Pioglitazone significantly reduced *HIF2 $\alpha$*  induction by 2.2-fold ( $P = 0.027$ ) and fully counteracted *CITED2* induction. Pioglitazone alone has no effect compared to control. Fig. 4b, According to IS, relapse was declared 2 consecutive positives, 6 months apart. For Extended Data Figures, statistical information is included in their legends.

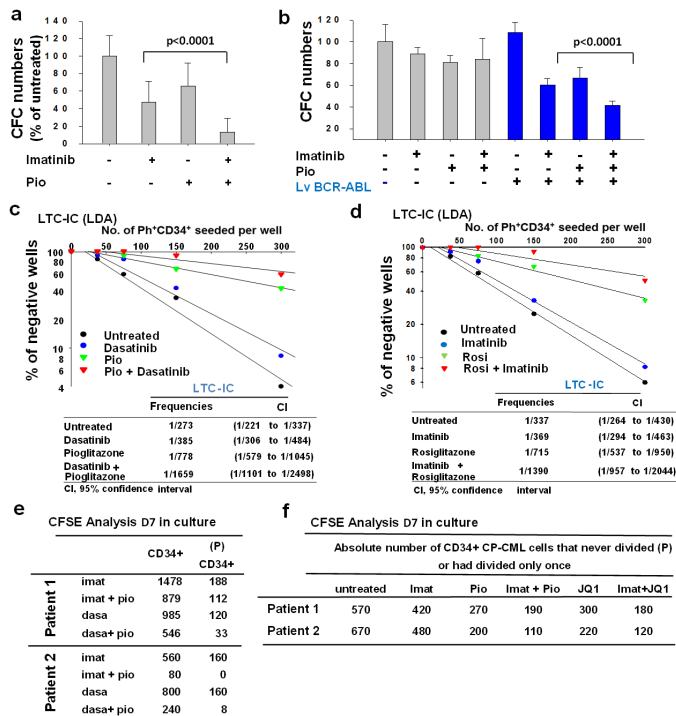
**Statistical information regarding synergy determination.** The putative synergistic effect of multiple drugs was determined by the algorithm and definitions of the Chou-Talalay medium-effect method<sup>36</sup>. The imatinib concentration required for 50% inhibition of the number of colonies obtained after CFC assay,  $IC_{50}^{imatinib}$ , was first determined in CD34<sup>+</sup> cells from 5 CP-CML patients at diagnosis. There was marked variability between patients ( $0.6 \mu M < IC_{50} < 2 \mu M$  with a median at  $1 \mu M$ ). Accordingly, at  $1 \mu M$  imatinib concentration, the percentage inhibition was  $47.1\% \pm 24$  of untreated CFC numbers in our full cohort of 29 CP-CML patients (Extended Data Figure 1). Percentages of inhibition were  $65.7\% \pm 26$  and  $12\% \pm 15$  of untreated CFC numbers for pioglitazone alone ( $10 \mu M$ ) and combination (imatinib  $1 \mu M$ , pioglitazone  $10 \mu M$ ), respectively. A combination index (CI)  $< 1$  defined synergy. We calculated  $IC_{50}$  and CI values with the Calcsyn software (Biosoft, Cambridge, UK). Imatinib and pioglitazone were assumed as having independent modes of action. In these conditions, CI were always less than 0.248, thus indicating synergy between the two drugs.

26. Davies, G. F., Juurlink, B. H. & Harkness, T. A. Troglitazone reverses the multiple drug resistance phenotype in cancer cells. *Drug Des. Devel. Ther.* **3**, 79–88 (2009).
27. Yuan, H. et al. Activation of stress response gene *SIRT1* by BCR-ABL promotes leukemogenesis. *Blood* **119**, 1904–1914 (2012).
28. Pitulis, N., Papageorgiou, E., Tenta, R., Lembessis, P. & Koutsilieris, M. IL-6 and *PPAR $\gamma$*  signalling in human PC-3 prostate cancer cells. *Anticancer Res.* **29**, 2331–2337 (2009).
29. Chen, Y., Hu, Y., Zhang, H., Peng, C. & Li, S. Loss of the *Alox5* gene impairs leukemia stem cells and prevents chronic myeloid leukemia. *Nature Genet.* **41**, 783–792 (2009).
30. Kominsky, D. J. et al. Abnormalities in glucose uptake and metabolism in imatinib-resistant human BCR-ABL-positive cells. *Clin. Canc. Res.* **15**, 3442–3450 (2009).
31. Lu, D. & Carson, D. A. Repression of beta-catenin signaling by *PPAR* gamma ligands. *Eur. J. Pharmacol.* **636**, 198–202 (2010).
32. Ito, K. et al. A PML-PPAR- $\delta$  pathway for fatty acid oxidation regulates hematopoietic stem cell maintenance. *Nature Med.* **18**, 1350–1358 (2012).
33. Gabert, J. et al. Standardization and quality control studies of 'real-time' quantitative reverse transcriptase polymerase chain reaction of fusion gene transcripts for residual disease detection in leukemia — a Europe Against Cancer program. *Leukemia* **17**, 2318–2357 (2003).
34. Nègre, D. et al. Characterization of novel safe lentiviral vectors derived from simian immunodeficiency virus (SIVmac251) that efficiently transduce mature human dendritic cells. *Gene Ther.* **7**, 1613–1623 (2000).
35. Roth, O. et al. Imatinib assay by HPLC with photodiode-array UV detection in plasma from patients with chronic myeloid leukemia: comparison with LC-MS/MS. *Clin. Chim. acta.* **411**, 140–146 (2010).
36. Chou, T. C. Drug combination studies and their synergy quantification using the Chou-Talalay method. *Cancer Res.* **70**, 440–446 (2010).



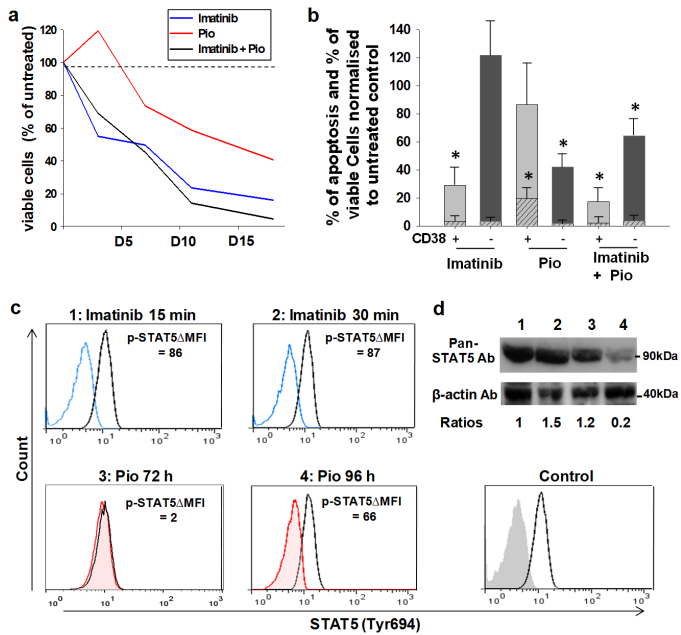


**Extended Data Figure 1 | Clonogenicity assays in the presence of various PPAR $\gamma$  agonists and validation of STAT5B overexpression and anti-PPAR $\gamma$ , anti-STAT5 and anti-HIF2 $\alpha$  siRNA.** **a**, Clonogenic capacities of BM CD34 $^{+}$  cells were assayed following pre-incubation for 2 days with culture medium alone (control) or supplemented with PPAR $\gamma$  agonists, PGJ $_2$ , troglitazone (Tro), ciglitazone (Cig), rosiglitazone (Ros), pioglitazone (Pio) or MCC-555 (MCC) (25  $\mu$ M each) (samples from 4 donors in triplicate). The number of colonies scored is expressed as percentage of control (untreated) values with standard deviation (s.d.), \* $P < 0.05$  using the nonparametric Wilcoxon rank test. **b**, Validation of anti-PPAR $\gamma$  siRNA used in Fig. 2b. CD34 $^{+}$  cells were transfected with irrelevant or PPAR $\gamma$  targeting siRNA (25 nM each). An anti-PPAR $\gamma$  shRNA was used as a positive control. PPAR $\gamma$  transcripts were normalized to GAPDH transcripts and expressed relative to the levels measured in untransfected cells. **c**, Western blot analysis with PPAR $\gamma$ , pan-STAT5, HIF2 $\alpha$  and anti-actin antibodies (Ab). Validation of siRNA against PPAR $\gamma$  or STAT5 and lentivector expressing STAT5B (LvSTAT5B) were realized on CD34 $^{+}$  cells from human UCB. Validation of siRNA against HIF2 $\alpha$  was realized on K562 cell line. Ctrl, scrambled siRNA; -, untreated. Quantification of western blot signal was realized with ImageJ software (<http://rsb.info.nih.gov/ij/>). Histograms show mean values with s.d.,  $n = 3$ .

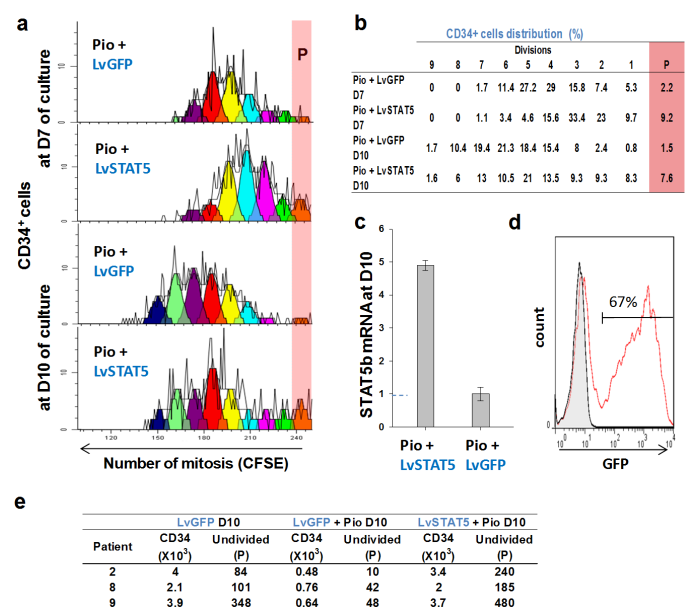


### Extended Data Figure 2 | Differential and synergistic effects of pioglitazone and TKIs on CML cells.

**a**, CFC assays with CD34<sup>+</sup> CP-CML cells from patients at diagnosis. Imatinib and/or pioglitazone were added for 48h before CFC assays. Means of 29 patients with standard deviation (s.d.). **b**, CFC assays after lentivector-mediated expression of BCR-ABL or eGFP (negative control) in human cord blood CD34<sup>+</sup> cells. Imatinib and/or pioglitazone were added for 48 h before CFC assays. Means of 3 individuals in triplicate with s.d. **c**, **d**, Limited dilution analysis (LDA) of CML LSCs by LTC-IC and frequency analysis. Plotted are means for CD34<sup>+</sup> cells from 2 CP-CML patients, 16 replica each. Imatinib 1  $\mu$ M, Rosi 10  $\mu$ M. **e**, **f**, CFSE analysis of CD34<sup>+</sup> cells (>96% Ph<sup>+</sup>) from CP-CML patients (for all experiments, imatinib 1  $\mu$ M, dasatinib 0.146  $\mu$ M, pioglitazone and rosiglitazone 10  $\mu$ M, JQ1 1  $\mu$ M. imat, imatinib; dasa, dasatinib; pio, pioglitazone; (P), undivided). To confirm the pivotal role played by STAT5 in the mechanism of action of pioglitazone in eroding the pool of TKI-resistant CML-LSCs, we investigated here the effect of the bromodomain inhibitor JQ1, which inhibits the transcriptional function of STAT5 by decreasing its activity through targeting the bromodomain-containing protein 2 (BRD2), a key cofactor of STAT5. Although this study with JQ1 is corroborative, one cannot completely exclude the possibility that these effects are coincidental, as targeting BRDs may cause a series of effects independent of STAT5.

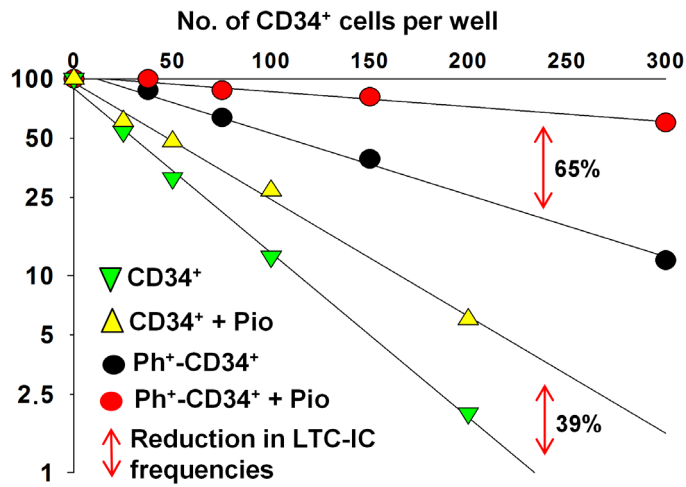


**Extended Data Figure 3 | Pioglitazone slowly decreases STAT5 expression whereas imatinib rapidly inhibits STAT5 phosphorylation.** **a**, Differential kinetics of action of imatinib and pioglitazone. CD34<sup>+</sup> CP-CML cells (patient 4) in liquid culture in serum-free medium without cytokines. **b**, Rate of apoptosis in CP-CML cell populations after 4 days of culture with imatinib and/or pioglitazone ( $n = 5$ ;  $*P < 0.05$ ). Solid bars (black for CD38<sup>-</sup> and grey for CD38<sup>+</sup>), percentage of recovery relative to input and normalized to untreated controls. Hatched bars, percentage of apoptosis, defined by the expression of annexin V. **c**, Flow cytometry analysis of permeabilized K562 cells with IgG against phosphorylated (Tyr694) STAT5. Untreated (black) and drug treated (red or blue). Control panel, no drug treatment but irrelevant IgG isotype control (grey peak). **d**, Western blot analysis with pan-STAT5 and anti-actin antibodies, showing a decrease of STAT5 by 3.5 fold  $\pm$  0.5 (s.d.) in lane 4 ( $n = 3$ ). Lanes 1 and 2 for imatinib (15 and 30 min exposure, respectively); lanes 3 and 4 for Pio (72 and 96 h exposure, respectively). Ratio indicates ratio of STAT5 expression/ $\beta$ -actin expression relative to lane 1. Quantification of western blot signals ( $n = 3$  for each condition) was realized with ImageJ software (<http://rsb.info.nih.gov/ij/>).

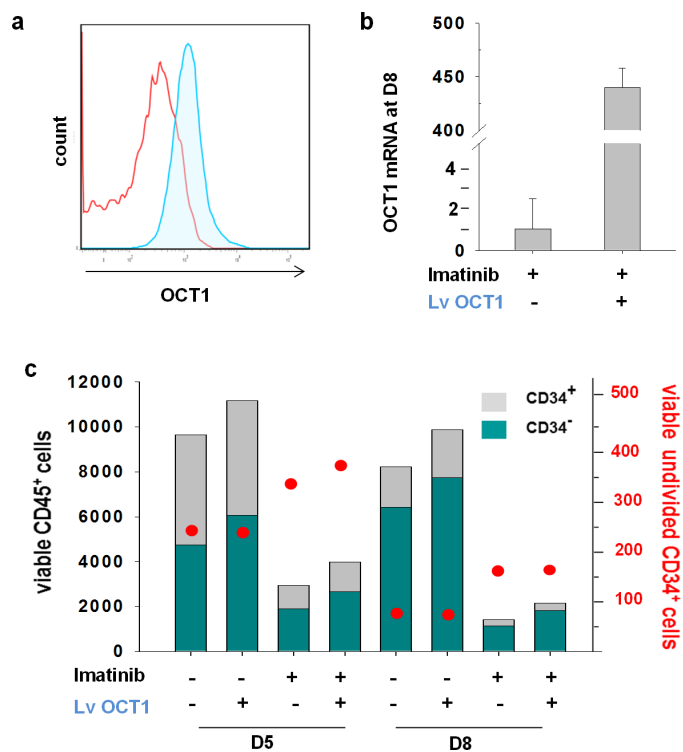


**Extended Data Figure 4 | Forced expression of STAT5 in CP-CML CD34<sup>+</sup> cells increases the compartment of quiescent cells.** **a**, CFSE analysis of CP-CML CD34<sup>+</sup> cells treated with pioglitazone after transduction with lentivectors (Lv) expressing eGFP or STAT5B, whose transcription is PPAR $\gamma$ -independent. Representative CP-CML patient 2 in triplicate (data for all patients are in Extended Data Fig. 2e). One coloured peak for each cell division number. P, colcemid arrested “parent-cells”. **b**, Distribution (%) of CD34<sup>+</sup> cells in each division peak shown in Extended Data Fig. 3a. **c**, *STAT5* mRNA expression analysis. **d**, Transduction efficiency of STAT5 lentivector. (5 replica with s.d.). **e**, Data for the 3 patients tested.

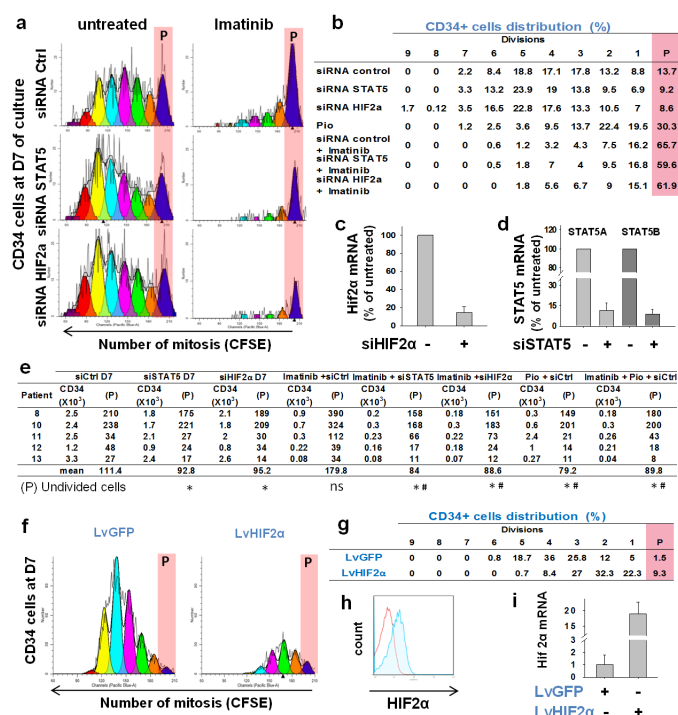




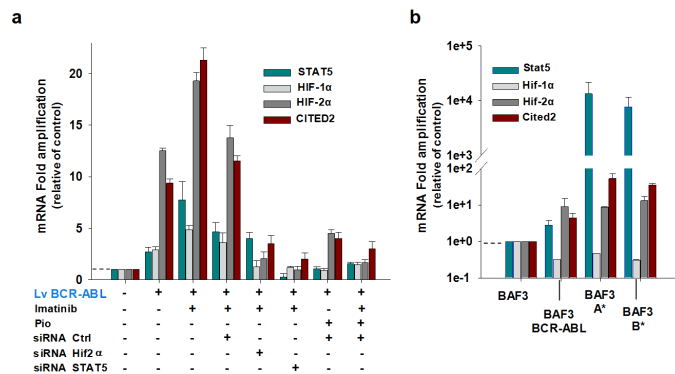
Extended Data Figure 5 | High toxicity of pioglitazone for CML LSCs vs. low toxicity for normal HSCs. LTC-IC (LDA) showing differential toxicity of pioglitazone for CP-CML vs. normal CD34<sup>+</sup> cells ( $n = 3, 16$  replica for each).



**Extended Data Figure 6 | Erosion of undivided and imatinib-resistant CD34<sup>+</sup> CP-CML cells is OCT1-independent.** **a**, Efficiency of LvOCT1 transduction (D8). **b**, *OCT1* mRNA expression. Results are normalized to *GAPDH* mRNA levels and represented relative to mRNA expression for the “imatinib alone” condition. **c**, CFSE analysis and absolute cell count in the presence of imatinib, with or without OCT1 overexpression. Left scale (black), total cells showing CD34<sup>+</sup> vs. CD34<sup>-</sup> cells (histograms). Right scale (red), undivided CD34<sup>+</sup> cells (red dots) (representative for  $n = 3$  CP-CML patients).

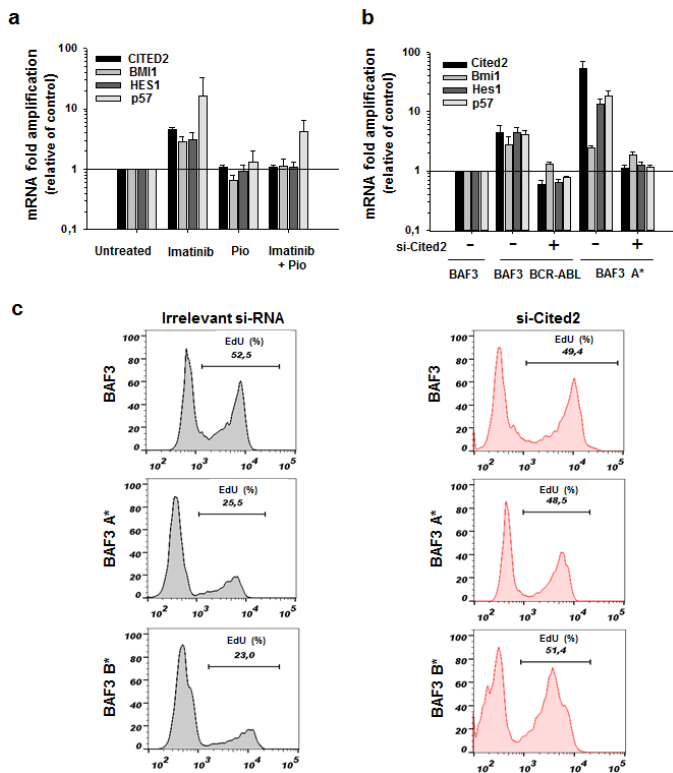


**Extended Data Figure 7 | The viability of undivided (P) and imatinib-resistant CD34<sup>+</sup> CP-CML cells depend on HIF2α expression.** Representative CP-CML patient 8 in triplicate (data for all patients are in Extended Data Fig. 6e). **a**, CFSE analysis in presence of siRNA against STAT5 or HIF2α in CD34<sup>+</sup>-Ph<sup>+</sup> cells treated or not with imatinib. One colored peak for each cell division number. P, colchimid arrested 'parent-cells'. **b**, Distribution (%) of CD34<sup>+</sup> cells in each division peak. **c**, HIF2α mRNA expression 72 h after siHIF2α transfection. **d**, STAT5 A and B mRNA expression 72 h after siSTAT5 transfection into human UCB CD34<sup>+</sup> cells. Results are normalized to GAPDH mRNA levels (means of 5 experiments with s.d. for each gene assessed). **e**, Data for the 5 patients tested (\**P* < 0.05 relative to siCtrl; #*P* < 0.05 relative to Imatinib + siCtrl). **f**, CFSE analysis of cord blood CD34<sup>+</sup> cells after transduction with lentivectors (Lv) expressing HIF2α or eGFP. One coloured peak for each cell division number. P, colchimid arrested 'parent-cells'. **g**, Distribution (%) of CD34<sup>+</sup> cells in each division peak (*n* = 5). **h**, Transduction efficiency of HIF2α Lv. **i**, HIF2α mRNA expression (means of 5 experiments with s.d.).



**Extended Data Figure 8 | Expression of target genes in CD34<sup>+</sup> cells and Ba/F3 cell line CML-models.** **a**, mRNA expression of target genes in CD34<sup>+</sup> cells from UCB transduced or not by BCR-ABL expressing lentivector (Lv). BCR-ABL<sup>+</sup> cells were cultured in serum-free medium without cytokines for 7 days with either imatinib alone (1  $\mu$ M) or imatinib and pioglitazone (1  $\mu$ M and 10  $\mu$ M, respectively) (means of 5 experiments with s.d. for each gene assessed). Results are normalized to *GAPDH* mRNA levels and represented relative to mRNA expression for the 'untreated' condition. Overexpression of BCR-ABL in CD34<sup>+</sup> cells from umbilical cord blood induced expression of *STAT5* and *HIF1 $\alpha$*  mRNAs by 2.7- and 2.8-fold, respectively ( $P = 0.043$ ), while *HIF2 $\alpha$*  and *CITED2* mRNAs were increased by 12.5-fold and 9-fold ( $P = 0.043$ ), respectively. In the presence of imatinib, *STAT5* and *HIF1 $\alpha$*  mRNAs were increased by 7.5- and 4.9-fold ( $P = 0.043$ ), respectively, while *HIF2 $\alpha$*  and *CITED2* were increased by 19.3- and 22-fold ( $P = 0.043$ ), respectively. Either pioglitazone or an siRNA against *STAT5* (A and B) significantly reduced the levels of *HIF2 $\alpha$*  and *CITED2* mRNAs, while an siRNA against *HIF2 $\alpha$*  significantly reduced *CITED2* mRNA expression (>threefold each,  $P < 0.05$ ). **b**, mRNA expression of target genes in Ba/F3 cell sub-lines independent of IL3 for viability after transduction with LvBCR-ABL or constitutively activated Stat5A1\*6 (A\*) or Stat5B1\*6 (B\*). Results are normalized to *GAPDH* mRNA levels and represented relative to mRNA expression for the original Ba/F3 cell line (means of 5 experiments with s.d. for each gene assessed). Forced expression of BCR-ABL increased the level of murine endogenous *Stat5* (a and b) mRNAs by 2.7 fold ( $P = 0.043$ ). When BCR-ABL or constitutively activated murine *Stat5* 1\*6 (a or b) were overexpressed, murine endogenous *Hif1 $\alpha$*  mRNA level was decreased by threefold ( $P = 0.043$ ) and murine endogenous *Hif2 $\alpha$*  and *Cited2* mRNAs increased by more than eightfold each ( $P = 0.043$ ).





**Extended Data Figure 9 | The key regulator of HSC quiescence, *CITED2*, is overexpressed in TKI-resistant CD34<sup>+</sup> cells from CP-CML patients.**

**a**, mRNA expression of *CITED2* and target genes thereof *BMI1*, *HES1* and *p57* after 9 days of culture with or without imatinib and pioglitazone. Results are normalized to GAPDH ( $n = 4$ ). **b**, mRNA expression of endogenous murine *Cited2* and its target genes *Bmi1*, *Hes1* and *p57* in Ba/F3 cell line with or without forced expression of BCR-ABL or constitutively active Stat5A 1\*6 (A\*) in the presence or not of siRNA against *Cited2*. Results are normalized to GAPDH (mean  $\pm$  s.d. of 3 independent experiments in triplicate). Forced expression of a constitutively active form of murine Stat5 1\*6 (A or B) in Ba/F3 cells, in and of itself, was sufficient to increase endogenous expression of murine *Cited2* markedly (52-fold) as well as that of its target genes *Bmi1* (2.5-fold), *Hes1* (13-fold) and *p57* (18-fold) **c**, Proliferation analysis by EdU incorporation assay of the Ba/F3 cell line that expresses or not constitutively active forms of Stat5 A 1\*6 (A\*) or B 1\*6 (B\*) in the presence or not of siRNA against *Cited2* (representative result of 5 independent experiments).

Extended Data Table 1 | CFSE analysis of CD34<sup>+</sup> cells (>96% Ph<sup>+</sup>) from 6 CP-CML patients after liquid culture without cytokines

Patient	Untreated D10		Imatinib D10		Pio D10		Imatinib + Pio D10	
	CD34 ( $\times 10^3$ )	Undivided (P)	CD34 ( $\times 10^3$ )	Undivided (P)	CD34 ( $\times 10^3$ )	Undivided (P)	CD34 ( $\times 10^3$ )	Undivided (P)
2	13	310	4.0	410	6.0	60	0.8	45
3	3.6	44	0.6	60	2.8	22	0.5	15
4	15	129	3.6	150	6.0	69	2.3	27
5	3.0	99	1.0	132	0.7	60	0.5	81
6	15	128	2.4	63	9.0	35	1.3	48
7	7.8	220	1.3	251	1.8	153	0.8	107
mean	155		177.6		66.5		53.8	
				P=0.248		P=0.027		P=0.027
						P=0.027		P=0.027
								P=0.345

# The spliceosome is a therapeutic vulnerability in MYC-driven cancer

Tiffany Y.-T. Hsu<sup>1,2,3,4</sup>, Lukas M. Simon<sup>4</sup>, Nicholas J. Neill<sup>1,4</sup>, Richard Marcotte<sup>5</sup>, Azin Sayad<sup>5</sup>, Christopher S. Bland<sup>1,4</sup>, Gloria V. Echeverria<sup>6,7,8</sup>, Tingting Sun<sup>1,4</sup>, Sarah J. Kurley<sup>1,4</sup>, Siddhartha Tyagi<sup>1,4</sup>, Kristen L. Karlin<sup>1,4</sup>, Rocio Dominguez-Vidaña<sup>1,2,4</sup>, Jessica D. Hartman<sup>4</sup>, Alexander Renwick<sup>4</sup>, Kathleen Scorsone<sup>9</sup>, Ronald J. Bernardi<sup>9</sup>, Samuel O. Skinner<sup>1,10</sup>, Antrix Jain<sup>1</sup>, Mayra Orellana<sup>1,4</sup>, Chandraiah Lagisetty<sup>11</sup>, Ido Golding<sup>1,10</sup>, Sung Y. Jung<sup>1</sup>, Joel R. Neilson<sup>2,6</sup>, Xiang H.-F. Zhang<sup>12</sup>, Thomas A. Cooper<sup>6,7,8</sup>, Thomas R. Webb<sup>11</sup>, Benjamin G. Neel<sup>5,13</sup>, Chad A. Shaw<sup>4</sup> & Thomas F. Westbrook<sup>1,2,4</sup>

MYC (also known as c-MYC) overexpression or hyperactivation is one of the most common drivers of human cancer. Despite intensive study, the MYC oncogene remains recalcitrant to therapeutic inhibition. MYC is a transcription factor, and many of its pro-tumorigenic functions have been attributed to its ability to regulate gene expression programs<sup>1–3</sup>. Notably, oncogenic MYC activation has also been shown to increase total RNA and protein production in many tissue and disease contexts<sup>4–7</sup>. While such increases in RNA and protein production may endow cancer cells with pro-tumour hallmarks, this increase in synthesis may also generate new or heightened burden on MYC-driven cancer cells to process these macromolecules properly<sup>8</sup>. Here we discover that the spliceosome is a new target of oncogenic stress in MYC-driven cancers. We identify *BUD31* as a MYC-synthetic lethal gene in human mammary epithelial cells, and demonstrate that *BUD31* is a component of the core spliceosome required for its assembly and catalytic activity. Core spliceosomal factors (such as SF3B1 and U2AF1) associated with *BUD31* are also required to tolerate oncogenic MYC. Notably, MYC hyperactivation induces an increase in total precursor messenger RNA synthesis, suggesting an increased burden on the core spliceosome to process pre-mRNA. In contrast to normal cells, partial inhibition of the spliceosome in MYC-hyperactivated cells leads to global intron retention, widespread defects in pre-mRNA maturation, and deregulation of many essential cell processes. Notably, genetic or pharmacological inhibition of the spliceosome *in vivo* impairs survival, tumorigenicity and metastatic proclivity of MYC-dependent breast cancers. Collectively, these data suggest that oncogenic MYC confers a collateral stress on splicing, and that components of the spliceosome may be therapeutic entry points for aggressive MYC-driven cancers.

To discover genes and cellular processes required to tolerate oncogenic MYC expression, we previously performed a genome-wide MYC-synthetic lethal screen in human mammary epithelial cells (HMECs) engineered with an inducible MYC and oestrogen receptor fusion protein (MYC-ER) for candidates affecting cell viability in a MYC-selective manner<sup>9</sup>. This screen nominated *BUD31* as a candidate MYC-synthetic lethal gene (Fig. 1a), in which barcoded *BUD31* short hairpin RNAs (shRNAs) consistently dropped out of the population in MYC-hyperactivated cells relative to cells without MYC induction (Fig. 1b). In validation experiments, *BUD31* depletion restrained clonogenic growth and activated apoptosis in MYC-induced cells, as compared to MYC-normal cells (Extended Data Fig. 1a–c). Expression of shRNA-resistant *BUD31* rescued the MYC-synthetic lethal phenotype

of *BUD31* shRNA (Fig. 1c and Extended Data Fig. 1d), indicating that the phenotype is an RNA interference (RNAi) on-target effect.

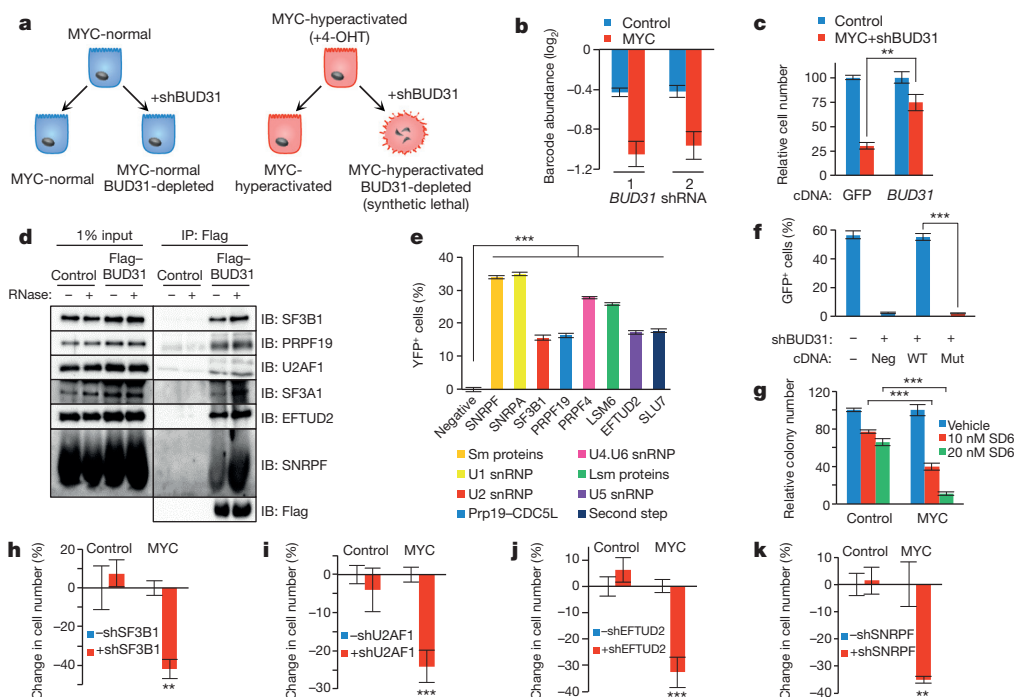
*BUD31* has been linked to the spliceosome in yeast<sup>10</sup>, but its function in mammalian systems has not been determined. To uncover the molecular function(s) of *BUD31*, we identified *BUD31*-interacting proteins by Flag-tagged *BUD31* immunoprecipitation from cells with or without RNase A (which eliminates protein–protein interactions mediated by RNA tethering), followed by mass spectrometry. Remarkably, 79 out of 134 core spliceosomal components were associated with *BUD31* (Extended Data Fig. 2a), suggesting a strong association between *BUD31* and the spliceosome in human cells.

The spliceosome is a dynamic molecular machine consisting of several nuclear protein complexes that cycle on and off of pre-mRNA during intronic splicing<sup>11</sup>. Co-immunoprecipitation experiments confirmed that *BUD31* associates with several subcomplexes of the spliceosome, including the Prp19–CDC5L subcomplex (PRPF19), the U2 small nuclear ribonucleoprotein particles (snRNPs; SF3B1 and SF3A1), U2-related factors (U2AF1), the U5 snRNP (EFTUD2), and Sm proteins (SNRPF) (Fig. 1d and Extended Data Fig. 2c), but interaction with non-spliceosomal proteins was not detected (Extended Data Fig. 2d, e). To test more broadly the association of *BUD31* with subcomplexes of the spliceosome, we performed bimolecular fluorescence complementation (BiFC) between *BUD31* and proteins from each major spliceosomal subcomplex. BiFC analysis indicated that *BUD31* associates with components of the major snRNPs (U1, U2, U4/U6 and U5) as well as Sm proteins (Fig. 1e and Extended Data Fig. 2b), indicating that *BUD31* is present at several stages of spliceosomal assembly.

To examine more directly whether *BUD31* has a role in pre-mRNA splicing, we tested *in vitro* splicing efficiency using nuclear extracts with or without *BUD31* knockdown. *BUD31* loss significantly inhibited pre-mRNA splicing (Extended Data Fig. 2f–i). In addition, knockdown of *BUD31* led to defects in early spliceosome assembly, as indicated by impaired formation of complex A (Extended Data Fig. 2h, i). Collectively, these data indicate that HMECs require a core spliceosomal protein (*BUD31*) to tolerate dysregulated MYC.

We proposed that cells with oncogenic MYC required *BUD31* for cell survival because of its role in the spliceosome. To test this hypothesis, we generated a *BUD31* mutant deficient in binding core spliceosomal proteins by mutating a highly conserved region spanning a C<sub>2</sub>–C<sub>2</sub> zinc-finger. Mutation of this region abrogated *BUD31* interaction with spliceosomal proteins (Extended Data Fig. 2j). To determine whether this region is also necessary for cells to tolerate MYC hyperactivation, we performed an *in vitro* competition assay. Green

<sup>1</sup>Verna & Marrs McLean Department of Biochemistry and Molecular Biology, Baylor College of Medicine, Houston, Texas 77030, USA. <sup>2</sup>Interdepartmental Program in Molecular and Biomedical Sciences, Baylor College of Medicine, Houston, Texas 77030, USA. <sup>3</sup>Medical Scientist Training Program, Baylor College of Medicine, Houston, Texas 77030, USA. <sup>4</sup>Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas 77030, USA. <sup>5</sup>Princess Margaret Cancer Centre, University Health Network, Toronto M5G 2C4, Canada. <sup>6</sup>Department of Molecular Physiology and Biophysics, Baylor College of Medicine, Houston, Texas 77030, USA. <sup>7</sup>Department of Pathology and Immunology, Baylor College of Medicine, Houston, Texas 77030, USA. <sup>8</sup>Department of Molecular and Cellular Biology, Baylor College of Medicine, Houston, Texas 77030, USA. <sup>9</sup>Department of Pediatrics, Baylor College of Medicine, Houston, Texas 77030, USA. <sup>10</sup>Department of Physics, University of Illinois, Urbana, Illinois 61801, USA. <sup>11</sup>Center for Chemical Biology, Bioscience Division, SRI International, Menlo Park, California 94025, USA. <sup>12</sup>The Lester and Sue Smith Breast Center, Baylor College of Medicine, Houston, Texas 77030, USA. <sup>13</sup>Department of Medical Biophysics, University of Toronto, Toronto M5S 2J7, Canada. †Present address: Humacyte, Morrisville, North Carolina 27560, USA.



**Figure 1 | The spliceosome is required for cells to tolerate oncogenic MYC hyperactivation.** **a**, *BUD31* is a MYC-synthetic lethal gene. **b**, *BUD31* shRNA (shBUD31) barcode abundances with/without MYC-ER hyperactivation (mean  $\pm$  s.e.m.,  $n = 3$  biological replicates). **c**, Relative number of MYC-ER HMECs with dox-inducible shRNA targeting the 3' untranslated region (UTR) of *BUD31*, and constitutive shRNA-resistant Flag-GFP or Flag-BUD31 expression (mean  $\pm$  s.e.m.,  $n = 4$  technical replicates). **d**, Flag-BUD31 co-immunoprecipitation for core spliceosomal factors. **e**, Interaction between *BUD31* and spliceosomal proteins assessed by BiFC (mean  $\pm$  s.e.m.,  $n = 3$

technical replicates). **f**, GFP<sup>+</sup> MYC-dependent cells with inducible shBUD31-UTR and constitutive wild-type, mutant *BUD31*, or negative control cDNA expression were mixed with GFP<sup>+</sup> cells and passaged (mean  $\pm$  s.e.m.,  $n = 8$  technical replicates, two-tailed Student's *t*-test). **g**, Change in MYC-ER HMEC clonogenicity after SD6 treatment (mean  $\pm$  s.e.m.,  $n = 4$  technical replicates, two-tailed Student's *t*-test). **h–k**, Relative number of MYC-ER HMECs after partial depletion of core spliceosomal proteins (mean  $\pm$  s.e.m.,  $n = 4$  technical replicates, one-way analysis of variance (ANOVA)). \*\* $P < 0.01$ , \*\*\* $P < 0.001$ .

fluorescent protein (GFP)-expressing MYC-driven breast cancer cells encoding inducible *BUD31* shRNA were transduced with shRNA-resistant wild-type or mutant *BUD31* complementary DNA, and these cells were mixed with non-transduced, GFP-negative cells. *BUD31* knockdown significantly inhibited the proliferation of MYC-driven cancer cells. Proliferation was fully rescued by wild-type *BUD31* cDNA but not by a *BUD31* mutant deficient in spliceosomal binding (Fig. 1f), suggesting that *BUD31* association with the spliceosome is required to support the survival of MYC-hyperactivated cells. More broadly, these results indicate that oncogenic MYC may increase cellular dependency on spliceosome function. By contrast, ectopic expression of the oncogenes *HER2* (also known as *ERBB2*) and *EGFR* did not enhance the effects of *BUD31* depletion (Extended Data Fig. 3a, b), suggesting that the stress imposed by MYC on spliceosomal function is not a universal feature of the oncogenic state.

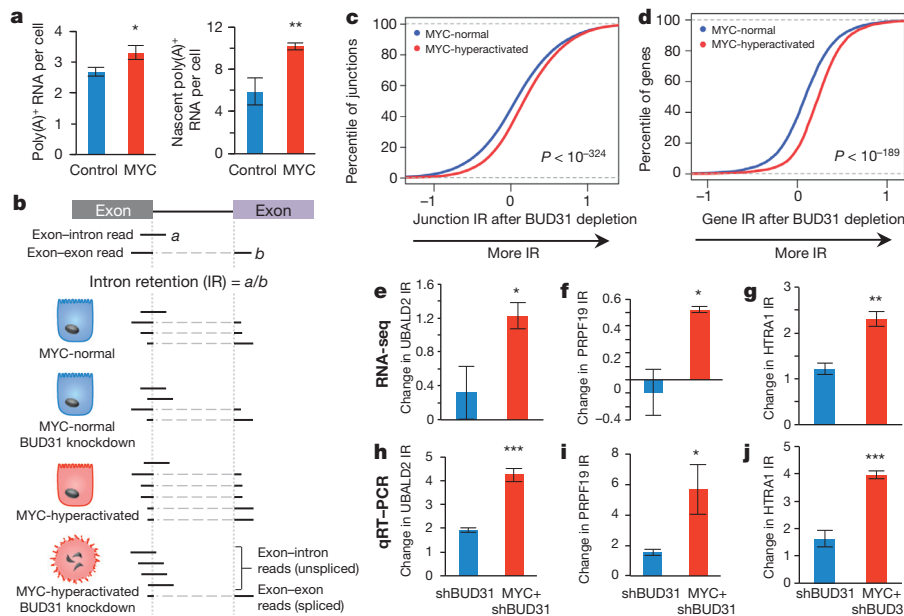
To test whether one or more subcomplexes of the spliceosome are required to tolerate aberrant MYC activity, we examined additional components of spliceosome assembly and catalysis including SF3B1 (U2 snRNP), U2AF1 (U2-related splicing factor), EFTUD2 (U5 snRNP) and SNRPF (core Sm protein found in every snRNP complex). Notably, partial depletion of each spliceosomal component led to loss of cell viability (Fig. 1h–k and Extended Data Fig. 4a–d) and increased apoptosis (Extended Data Fig. 4e–h) in MYC-hyperactivated cells. This suggests that several subcomplexes of the core spliceosome are required for cells to tolerate oncogenic MYC, and that MYC-hyperactivated cells are sensitive to modest perturbations in spliceosome function.

Next, we investigated whether pharmacological inhibition of the spliceosome is also synthetic lethal with MYC. Several pharmacological agents (for example, FR901464, pladienolides and their derivatives) have been characterized to bind the core SF3b spliceosomal complex components and inhibit spliceosome function<sup>12</sup>. However, most of these

inhibitors are not amenable for *in vivo* delivery. We developed a new small molecule inhibitor of SF3B1, known as SD6, that impairs spliceosome function and is bioavailable in mammals<sup>13</sup>. Consistent with our genetic data, low SD6 concentrations significantly suppressed colony formation (Fig. 1g) and induced apoptosis (Extended Data Fig. 4i) in a MYC-selective manner. The synthetic-lethal interaction between MYC hyperactivation and core spliceosome perturbation suggests that pre-mRNA splicing is necessary to tolerate oncogenic MYC.

In many different cell lineages and experimental systems, oncogenic MYC activation has been shown to amplify the synthesis of cellular mRNA through direct or indirect mechanisms<sup>4,5,14,15</sup>. In agreement, MYC hyperactivation in HMECs increased total cellular mRNA synthesis and mRNA steady-state levels (Fig. 2a) without an increase in cellular growth rate (Extended Data Fig. 3c). In contrast to a recent report in B-cell compartments<sup>16</sup>, MYC hyperactivation did not affect the levels of spliceosome proteins in HMECs (data not shown), suggesting that increased pre-mRNA dosage is not compensated for by higher spliceosome levels. Thus, we proposed that the MYC-induced increase in global mRNA synthesis confers increased pressure on the spliceosome to process pre-mRNAs, and partial perturbation of the spliceosome would lead to widespread defects in the splicing of pre-mRNA introns in the MYC-hyperactive state. To test this hypothesis, we compared intron retention (IR) after *BUD31* knockdown in MYC-normal or MYC-hyperactivated cells. We performed RNA-sequencing (RNA-seq) from cells in each state (normal, *BUD31* knockdown, MYC-hyperactive, and MYC-hyperactive with *BUD31* knockdown) and determined the pre-mRNA splicing efficiency by calculating IR at junctions across the genome (Fig. 2b). Because the analysis of intronic reads may be influenced by the presence of stable RNAs within introns and/or spliced lariats, we restricted the analysis to reads directly spanning exon–intron or exon–exon junction sequences (75,623 junctions in 6,861 genes) (see Methods).

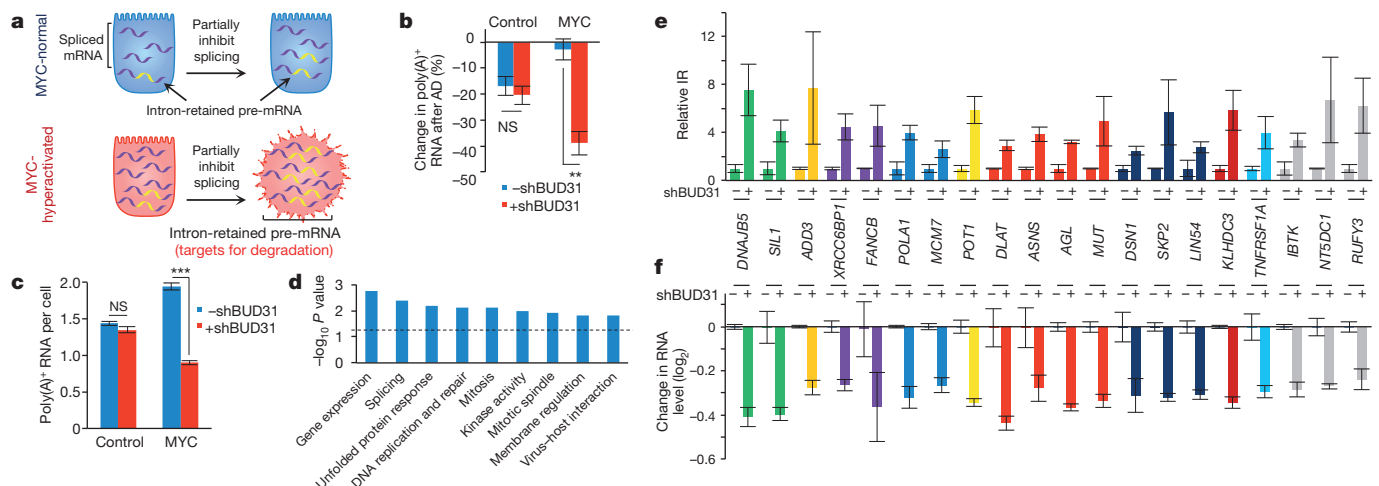




**Figure 2 | In MYC-hyperactivated cells, perturbation of the spliceosome leads to global intron retention.** **a**, Left, total poly(A)<sup>+</sup> RNA per cell ( $10^{-4}$  ng). Right, newly synthesized 4-sU-labelled poly(A)<sup>+</sup> RNA per cell ( $10^{-5}$  ng). Data are mean  $\pm$  s.e.m.,  $n = 4$  technical replicates for both assays, two-tailed Student's  $t$ -test. **b**, Schematic of IR analysis. **c**, **d**, Empirical cumulative distribution of IR coefficients for 75,623 exon-intron junctions (**c**) or 6,861 genes (**d**). Curves represent IR differences after BUD31 depletion in MYC-normal and MYC-hyperactive states. A rightward shift in the MYC-hyperactive curve indicates increased IR (Kolmogorov-Smirnov test). **e–g**, log<sub>2</sub>-fold changes in junction IR relative to untreated by RNA-seq of representative genes (mean  $\pm$  s.e.m.,  $n = 3$  biological replicates, two-tailed Student's  $t$ -test). **h–j**, qRT-PCR validation showing fold change in junction IR relative to untreated (mean  $\pm$  s.d.,  $n = 3$  biological replicates, two-tailed Student's  $t$ -test). \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$ .

To examine the effects of spliceosome perturbation in the normal and oncogenic MYC states, we compared the effect of BUD31 knockdown on junction IR coefficients in wild-type and MYC-hyperactivated cells. Notably, BUD31 depletion caused significantly more IR in the MYC-hyperactive state than in the MYC-normal state (Fig. 2c,  $P < 10^{-324}$ ). Similar results were observed when junction coefficients were computed on a gene level (Fig. 2d,  $P < 10^{-189}$ ). The increase in IR conferred by aberrant MYC activation and BUD31 shRNA was validated on individual exon-intron junctions via quantitative reverse transcriptase PCR (qRT-PCR) (examples in Fig. 2e–j). IR was not limited to a few discrete genes. Instead, BUD31 knockdown in the MYC-hyperactive state led to significantly increased IR in 42% of genes analysed (2,848 of 6,861,  $P < 0.05$ ). These data indicate that the combination of oncogenic MYC activation and partial spliceosome inhibition leads to a widespread increase in IR. This is consistent with the hypothesis that the MYC-induced increase in pre-mRNA synthesis enhances cellular dependency on optimal spliceosome function by raising the level of pre-mRNA substrates for spliceosomal processing.

Intron-retaining pre-mRNAs often fail to complete mRNA maturation and are commonly degraded via quality control mechanisms<sup>17</sup>. Because the combination of MYC hyperactivation and spliceosome inhibition led to a global increase in intron retention (Fig. 2c, d), we proposed that these cells may contain widespread defects in pre-mRNA maturation and stability (Fig. 3a). To test this hypothesis, we measured the levels of cellular poly(A)<sup>+</sup> RNA in each of the four states (with/without MYC hyperactivation, with/without BUD31 shRNA) before and after treatment with the transcriptional inhibitor actinomycin D. After actinomycin D treatment, cellular poly(A)<sup>+</sup> RNA decreased by comparable levels ( $\sim 16$ – $19\%$ ) in control cells with or without BUD31 knockdown (Fig. 3b). Notably, MYC-hyperactivated cells exhibited enhanced mRNA stability, perhaps resulting from increased polysomal loading of mRNA during MYC-induced translation<sup>18</sup>. By contrast, cells containing MYC hyperactivation and BUD31-depletion exhibited a substantially greater loss (38%) of poly(A)<sup>+</sup> RNA after actinomycin D treatment, suggesting a defect in pre-mRNA maturation and/or stability in the combined MYC-hyperactivated and BUD31-shRNA state. Similarly, fluorescence *in situ* hybridization (FISH) measurements



**Figure 3 | Combined spliceosomal perturbation and MYC hyperactivation inhibits pre-mRNA maturation.** **a**, Model of MYC-spliceosome synthetic lethality. **b**, Difference in cellular poly(A)<sup>+</sup> RNA in HMECs after actinomycin D (AD) treatment ( $n = 3$  biological replicates, two-tailed Student's  $t$ -test). **c**, Steady-state poly(A)<sup>+</sup> RNA levels per cell ( $10^{-4}$  ng) ( $n = 4$  biological replicates, two-tailed Student's  $t$ -test). **d**, Gene Ontology (GO) enrichment of

intron-retained genes in the MYC-hyperactive and BUD31-depleted state. Dashed line indicates  $P = 0.05$ . **e**, **f**, In MYC-hyperactive BUD31 shRNA cells, representative genes display increased IR (**e**) and decreased steady-state RNA levels (**f**) after BUD31 knockdown in MYC-hyperactivated cells. Bar colours represent GO terms, see legend in Extended Data Fig. 6. Data are mean  $\pm$  s.e.m. \*\* $P < 0.01$ , \*\*\* $P < 0.001$ . NS, not significant.

of poly(A)<sup>+</sup> RNA revealed that the combination of MYC hyperactivation and BUD31 knockdown led to a substantially greater decrease (60%) in poly(A)<sup>+</sup> RNA after actinomycin D treatment (Extended Data Fig. 5a). Similar trends were observed in nuclear RNA pools, consistent with defects in nuclear pre-mRNA maturation (Extended Data Fig. 5b). Consistent with this decrease in pre-mRNA maturation and stability, cells containing oncogenic MYC and BUD31 knockdown exhibited significantly lower (54%) steady-state levels of poly(A)<sup>+</sup> RNA (Fig. 3c). Collectively, these results indicate that MYC hyperactivation increases cellular pre-mRNA synthesis, and inhibition of the spliceosome reduces the cellular capacity to process this pre-mRNA burden. The result of this MYC-hyperactivated and spliceosome-hypomorphic state is enhanced intron retention, decreased mRNA maturation and stability, and a significant loss of steady-state cellular mRNA.

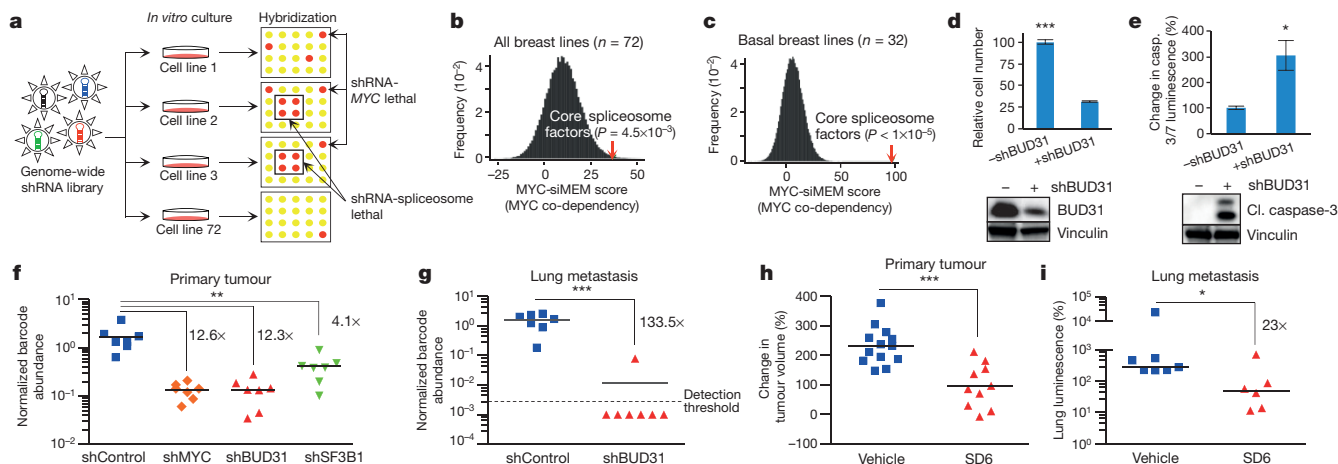
Gene Ontology analysis of genes with the most significant intron retention in the combined MYC-hyperactive and BUD31-depleted state (2,848 out of 6,816 genes analysed for IR) suggests that many essential processes and subcellular structures were affected, including gene expression, DNA replication and repair, the mitotic spindle, unfolded protein response, and RNA splicing (Fig. 3d). Many genes participating in these essential cell processes exhibited increased IR in the combined MYC-hyperactive and BUD31-knockdown state (representative genes in Fig. 3e) and a concomitant decrease in RNA levels, consistent with a defect in maturation and stability of IR-containing transcripts (Fig. 3f). Consistent with their role in crucial cellular processes, knockdown of these genes reduced cell number by 0.7–4.2-fold (as quantified by barcode-tag abundance, Extended Data Fig. 6). Together, these data are consistent with the hypothesis that the combination of oncogenic MYC and spliceosome inhibition leads to widespread loss of mRNA integrity, resulting in the deregulation of many essential genes and processes instead of a single pathway.

Because oncogenic MYC significantly increases the sensitivity of HMECs to inhibition of the spliceosome, we proposed that MYC-driven cancers may be hyperdependent on core spliceosomal function to support their survival. We queried whether MYC-driven breast cancer cell lines exhibit increased sensitivity to knockdown of core spliceosomal genes. Recently, we conducted genome-wide RNAi screens in a panel of 72 breast cancer and immortalized cell lines for genes affecting cell viability (Fig. 4a) (R.M., A.S. and B.G.N., manuscript in prepara-

tion). From this data set, we tested for a correlation between MYC-dependency (as indicated by sensitivity to MYC shRNAs) and dependency on the spliceosome (as indicated by sensitivity to shRNAs targeting spliceosome components in the shRNA library), or on 100,000 randomly drawn gene sets. Notably, MYC-dependent breast cancer cell lines were significantly more sensitive to shRNAs targeting the core spliceosome (Fig. 4b,  $P = 0.005$ ). The correlation between MYC-dependency and spliceosome-dependency was significantly pronounced in the basal breast cancer lines (Fig. 4c,  $P < 0.00001$ ), an aggressive molecular subtype of breast cancer frequently driven by MYC.

Triple-negative breast cancers are commonly driven by MYC, and exhibit an aggressive, highly metastatic clinical course. To determine whether MYC-driven triple-negative breast cancers are dependent on spliceosomal integrity for their tumorigenic and metastatic proclivity, we tested the effects of genetic and pharmacological inhibition of the spliceosome on MYC-dependent and metastatic triple-negative breast cancer (TNBC) models. Inducible BUD31 shRNA reduced cell viability and increased apoptosis in MYC-dependent TNBC cells *in vitro* (Fig. 4d, e and Extended Data Fig. 7a, b). Similar to MYC-ER HMECs, MYC protein levels remained unchanged during BUD31 depletion in these MYC-dependent cancer cell lines (Extended Data Fig. 8a, b), suggesting that the apoptotic response was not due to loss of the driver oncogene (MYC). To assess the effect of spliceosomal perturbation on tumour growth, we established a pooled competition assay that uses shRNA-associated barcodes to detect changes in tumour cell fitness (Extended Data Fig. 9). In the metastatic TNBC cell line MDA-MB-231-LM2 (LM2)<sup>19</sup>, inducible MYC-shRNA-expressing cells dropped out of the tumour population, confirming the MYC dependency of this TNBC model (Fig. 4f). Similarly, tumour cells containing BUD31 or SF3B1 shRNA dropped out of the tumour population (Fig. 4f). Tumorigenicity of another MYC-dependent TNBC model (SUM159) was similarly impaired by BUD31 depletion (Extended Data Fig. 7c, d). These data suggest that the loss of BUD31 or other core spliceosomal factors inhibits MYC-dependent breast cancer growth *in vivo*.

Because MYC-driven breast cancers are prone to metastasize to visceral organs including the lungs<sup>20</sup>, we tested whether perturbation of spliceosome function affected metastatic expansion of MYC-dependent LM2 cells. As shown in Fig. 4g, metastatic cells with BUD31 knockdown were significantly depleted from the population



**Figure 4 | *In vivo* perturbation of spliceosomal activity impairs MYC-dependent breast tumours and metastases.** **a**, Schematic for identifying genetic co-dependencies in breast cancer lines. **b**, **c**, MYC-siMEM (mixed-effect model) score, which represents the correlation between cell line sensitivity to MYC shRNAs and sensitivity to shRNAs targeting random gene sets ( $n = 100,000$ ; see Methods), is plotted against frequency of gene sets. Increasing MYC-siMEM values denote higher correlation with MYC-dependency. Red arrows indicate MYC-siMEM scores for spliceosome-dependency in all breast cancer lines ( $n = 72$ ) (**b**) and the basal breast cancer subset ( $n = 32$ ) (**c**).  $P$  value by bootstrap analysis for both. **d**, **e**, MDA-MB-231-LM2 cells with shBUD31 display diminished BUD31 protein levels (**d**, bottom), decreased cell

numbers (**d**, top) (mean  $\pm$  s.e.m.,  $n = 8$  technical replicates, two-tailed Student's  $t$ -test), and increased caspase-3 cleavage (**e**, bottom) and caspase-3/7 luminescence (**e**, top) (mean  $\pm$  s.e.m.,  $n = 3$  technical replicates, two-tailed Student's  $t$ -test). **f**, **g**, Barcode-shRNA abundance of LM2 cells within primary tumours (**f**) or pulmonary metastases (**g**). Mean barcode abundance in each tumour or lung is normalized to the injected cell population ( $n = 3$  technical replicates, two-tailed Student's  $t$ -test). **h**, Change in LM2 tumour growth after 2 weeks of vehicle ( $n = 13$ ) or SD6 ( $n = 10$ ) infusion. Bars indicate mean values (two-tailed Student's  $t$ -test). **i**, Pulmonary LM2 bioluminescence after 10-day infusion with vehicle ( $n = 7$ ) or SD6 ( $n = 6$ ). Bars indicate median values (Mann-Whitney test). \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$ .

(>133.5-fold change), with most doxycycline (dox)-positive tumours containing *BUD31* shRNA barcodes below the level of detection. These data suggest that *BUD31* and the spliceosome are essential for MYC-dependent breast tumorigenicity and metastatic expansion *in vivo*.

Next, we tested whether pharmacological inhibition of the spliceosome also impaired tumorigenic and metastatic potential of MYC-dependent TNBC cells. Compared to MYC-normal cell lines (half-maximal inhibitory concentration (IC<sub>50</sub>) value ≈ 53 nM), MYC-driven cancer cells were significantly more sensitive (IC<sub>50</sub> value ≈ 4 nM) to the spliceosome inhibitor SD6 *in vitro* (Extended Data Fig. 10a). Similarly, SD6 suppressed the proliferation of a MYC-driven B-cell model<sup>4</sup> (Extended Data Fig. 10b), suggesting that oncogenic MYC may confer hyperdependency on the spliceosome in many epigenetic backgrounds and cancer types. In primary LM2 tumour xenografts, SD6 potently restrained tumour growth with no toxicities in any organ system examined, suggesting that splicing is essential for the tumorigenicity of these MYC-dependent breast cancer cells (Fig. 4h). Similarly, SD6 impaired lung metastatic expansion in experimental metastasis assays (Fig. 4i), and extended progression-free survival (Extended Data Fig. 10c). Collectively, these data suggest that MYC-driven breast cancers depend on spliceosomal integrity for their tumorigenic and metastatic progression.

Altogether, the results suggest that MYC-driven breast cancers contain an enhanced dependency on the core spliceosome. Recent studies have shown that MYC regulates splicing of select genes via induction of alternative splicing factors or components of the core spliceosome<sup>16,21</sup>. This study suggests that MYC may induce a much broader stress on splicing via its ability to increase global pre-mRNA synthesis. Recently, there has been considerable investigation into how MYC elicits a widespread increase in mRNA synthesis across the transcriptome<sup>4,5,22,23</sup>. Notably, either direct or indirect mechanisms of increased pre-mRNA synthesis elicited by MYC could lead to an enhanced dependency on the spliceosome, and thus make MYC-driven cancers candidates for spliceosome-based therapies. These observations provoke the important question of whether MYC-induced amplification of mRNA synthesis may also generate vulnerabilities in other aspects of RNA processing (such as mRNA capping, polyadenylation or mRNA export) and downstream protein biosynthesis<sup>24</sup> in MYC-driven cancers. Notably, the spliceosome may be a target of both oncogene addiction and oncogenic stress. Components of the U2 snRNP, such as *SF3B1* and *U2AF1*, contain frequent and recurrent somatic mutations that cluster in an evolutionarily conserved domain, suggestive of oncogenic function<sup>25,26</sup>. On the basis of such putative oncogenic functions, the spliceosome has been proposed as a target for classical oncogene addiction, in which spliceosome mutant tumours may be addicted to the oncogenic functions of spliceosome mutants and thus sensitive to spliceosome inhibitors. However, this study and others<sup>27,28</sup> have shown that inhibition of spliceosome components is deleterious in cancer cell line models that lack spliceosome mutations, suggesting that other drivers of cancer (such as MYC) are determinants of sensitivity to spliceosome inhibitors. Because oncogenic MYC is known to drive several pro-tumorigenic programs that include rewiring of biosynthetic pathways<sup>29,30</sup>, our model provokes the important hypothesis that cellular processes (such as splicing) that enable cancer cells to tolerate such widespread shifts in macromolecular synthesis may provide entry points for anti-cancer therapies.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 27 August 2014; accepted 24 July 2015.

Published online 2 September 2015.

1. Eilers, M. & Eisenman, R. N. Myc's broad reach. *Genes Dev.* **22**, 2755–2766 (2008).
2. Sabo, A. & Amati, B. Genome recognition by MYC. *Cold Spring Harb. Perspect. Med.* **4**, a014191 (2014).
3. Dang, C. V. MYC, metabolism, cell growth, and tumorigenesis. *Cold Spring Harb. Perspect. Med.* **3**, a014217 (2013).
4. Lin, C. Y. *et al.* Transcriptional amplification in tumor cells with elevated c-Myc. *Cell* **151**, 56–67 (2012).

5. Nie, Z. *et al.* c-Myc is a universal amplifier of expressed genes in lymphocytes and embryonic stem cells. *Cell* **151**, 68–79 (2012).
6. Ruggero, D. The role of Myc-induced protein synthesis in cancer. *Cancer Res.* **69**, 8839–8843 (2009).
7. Barna, M. *et al.* Suppression of Myc oncogenic activity by ribosomal protein haploinsufficiency. *Nature* **456**, 971–975 (2008).
8. Luo, J., Solimini, N. L. & Elledge, S. J. Principles of cancer therapy: oncogene and non-oncogene addiction. *Cell* **136**, 823–837 (2009).
9. Kessler, J. D. *et al.* A SUMOylation-dependent transcriptional subprogram is required for Myc-driven tumorigenesis. *Science* **335**, 348–353 (2012).
10. Masciadri, B. *et al.* Characterization of the *BUD31* gene of *Saccharomyces cerevisiae*. *Biochem. Biophys. Res. Commun.* **320**, 1342–1350 (2004).
11. Wahl, M. C., Will, C. L. & Luhrmann, R. The spliceosome: design principles of a dynamic RNP machine. *Cell* **136**, 701–718 (2009).
12. Bonnal, S., Vigevani, L. & Valcarcel, J. The spliceosome as a target of novel antitumour drugs. *Nature Rev. Drug Discov.* **11**, 847–859 (2012).
13. Lagisetty, C. *et al.* Optimization of antitumor modulators of pre-mRNA splicing. *J. Med. Chem.* **56**, 10033–10044 (2013).
14. Kanazawa, S., Soucek, L., Evan, G., Okamoto, T. & Peterlin, B. M. c-Myc recruits P-TEFb for transcription, cellular proliferation and apoptosis. *Oncogene* **22**, 5707–5711 (2003).
15. Rahl, P. B. *et al.* c-Myc regulates transcriptional pause release. *Cell* **141**, 432–445 (2010).
16. Koh, C. M. *et al.* MYC regulates the core pre-mRNA splicing machinery as an essential step in lymphomagenesis. *Nature* **523**, 96–100 (2015).
17. Garneau, N. L., Wilusz, J. & Wilusz, C. J. The highways and byways of mRNA decay. *Nature Rev. Mol. Cell Biol.* **8**, 113–126 (2007).
18. Mezquita, P., Parghi, S. S., Brandvold, K. A. & Ruddell, A. Myc regulates VEGF production in B cells by stimulating initiation of VEGF mRNA translation. *Oncogene* **24**, 889–901 (2005).
19. Minn, A. J. *et al.* Distinct organ-specific metastatic potential of individual breast cancer cells and primary tumors. *J. Clin. Invest.* **115**, 44–55 (2005).
20. Di Cosimo, S. & Baselga, J. Management of breast cancer with targeted agents: importance of heterogeneity. *Nature Rev. Clin. Oncol.* **7**, 139–147 (2010).
21. David, C. J., Chen, M., Assanah, M., Canoll, P. & Manley, J. L. HnRNP proteins controlled by c-Myc deregulate pyruvate kinase mRNA splicing in cancer. *Nature* **463**, 364–368 (2010).
22. Sabó, A. *et al.* Selective transcriptional regulation by Myc in cellular growth control and lymphomagenesis. *Nature* **511**, 488–492 (2014).
23. Walz, S. *et al.* Activation and repression by oncogenic MYC shape tumour-specific gene expression profiles. *Nature* **511**, 483–487 (2014).
24. Lin, C. J. *et al.* Targeting synthetic lethal interactions between Myc and the eIF4F complex impedes tumorigenesis. *Cell Rep.* **1**, 325–333 (2012).
25. Graubert, T. A. *et al.* Recurrent mutations in the U2AF1 splicing factor in myelodysplastic syndromes. *Nature Genet.* **44**, 53–57 (2011).
26. Papaemmanuil, E. *et al.* Somatic *SF3B1* mutation in myelodysplasia with ring sideroblasts. *N. Engl. J. Med.* **365**, 1384–1395 (2011).
27. Hubert, C. G. *et al.* Genome-wide RNAi screens in human brain tumor isolates reveal a novel viability requirement for PHF5A. *Genes Dev.* **27**, 1032–1045 (2013).
28. Adler, A. S. *et al.* An integrative analysis of colon cancer identifies an essential function for PRPF6 in tumor growth. *Genes Dev.* **28**, 1068–1084 (2014).
29. Cunningham, J. T., Moreno, M. V., Lodi, A., Ronen, S. M. & Ruggero, D. Protein and nucleotide biosynthesis are coupled by a single rate-limiting enzyme, PRPS2, to drive cancer. *Cell* **157**, 1088–1103 (2014).
30. Liu, Y. C. *et al.* Global regulation of nucleotide biosynthetic genes by c-Myc. *PLoS ONE* **3**, e2722 (2008).

**Acknowledgements** We would like to thank J. Rosen, S. Butler, K. Neugebauer, M. Moore, S. Elledge, T. Davoli, members of T.F.W., C.A.S. and T.A.C. laboratories for comments, and P. Yu for bioinformatics support. The authors also acknowledge the joint participation by Adrienne Helis Melvin Medical Research Foundation through its direct engagement in the continuous active conduct of medical research in conjunction with Baylor College of Medicine for cancer research. The Dan L. Duncan Cancer Center Shared Resources was supported by the NCI P30CA125123 Center Grant and provided technical assistance including Cell-Based Assay Screening Service (D. Liu), Genomic and RNA Profiling Resource (L. White), Biostatistics & Informatics Shared Resource (S. Hilsenbeck), Cytometry and Cell Sorting (J. Sederstrom; P30 AI036211 and S10 RR024574), and the Proteomics and Metabolomics Core Facility (Cancer Prevention and Research Institute of Texas, RP12009). T.Y.-T.H. was supported by NIH pre-doctoral fellowship (NCI 1F30CA180447) and CPRIT training grant (RP101499). M.O. and R.J.B. were supported by The Gillson Longenbaugh Foundation. R.J.B. was supported by Alex's Lemonade Stand Foundation. T.F.W. was supported by CPRIT (RP120583), the Susan G. Komen for the Cure (KG090355), the NIH (1R01CA178039-01 and U54-CA149196) and the DOD Breast Cancer Research Program (BC120604).

**Author Contributions** T.Y.-T.H., N.J.N., R.M., C.S.B., G.V.E., T.S., S.J.K., S.T., K.L.K., J.D.H., K.S., R.J.B., S.O.S., A.J., C.L. and M.O. performed the experiments. L.M.S., A.S., R.D.-V., A.R. and C.A.S. performed statistical analyses. I.G., S.Y.J., J.R.N., X.H.-F.Z., T.A.C., T.R.W., B.G.N., C.A.S. and T.F.W. devised or supervised experiments. T.Y.-T.H. and T.F.W. wrote the manuscript.

**Author Information** RNA-seq data sets have been deposited in the NCBI Gene Expression Omnibus (GEO) under accession number GSE66182. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to T.F.W. (thomasw@bcm.edu).



## METHODS

**Vectors and virus production.** Commercially available pGIPZ shRNAs targeting *BUD31* (V2LHS\_47771 and V2LHS\_47770), *EFTUD2* (V2LHS\_28167), *SF3B1* (V3LHS\_397872), *SNRPF* (V2LHS\_276933) and *U2AF1* (V2LHS\_84677) were obtained from Open Biosystems. shRNAs targeting the 3' UTR region of *BUD31* were designed using the BiopredSI and RNAi Codex algorithms (shRNA sequence 5'-TGCTGTTGACAGTGAGCGCCGCTGTCTATCAGCTG TGATTTAGTGAAGCCACAGATGTAAATCAGCTGATAGACAGCGATG CTTACTGCCTCGGA-3'). For inducible RNAi experiments, shRNAs were sub-cloned into the pINDUCER dox-inducible lentiviral expression system<sup>31</sup>. Lentiviruses and retroviruses were produced by transiently transfecting shRNA or cDNA constructs using Mirus Bio TransIT transfection protocols into 293T cells and collecting viral supernatants 48 h after transfection.

**Cell culture.** HMECs expressing hTERT and inducible MYC-ER (MYC-ER HMECs), F7 epithelial cells and human mammary epithelial HME1 cells were cultured in mammary epithelial growth medium (MEGM, Lonza). 293T cells, HeLa cells and MDA-MB-231-LM2 human breast cancer cells were cultured in DMEM (Gibco) supplemented with 10% FBS. SUM159 human breast cancer cells were cultured in F12 (Gibco) media supplemented with 5% FBS, 10 mM HEPES (Gibco), 5 µg ml<sup>-1</sup> insulin (Invitrogen), and 1 µg ml<sup>-1</sup> hydrocortisone. The P493-6 human B-cell lymphoma cell line was cultured in RPMI-1640 supplemented with 10% FBS (Clonotech) and 1% GlutaMAX (Invitrogen). All cell lines were incubated at 37 °C and 5% CO<sub>2</sub>. Cell lines were obtained from ATCC, and all cell lines are tested yearly for mycoplasma contamination. Stable cell lines expressing shRNAs or cDNAs were generated by lentiviral or retroviral transduction in the presence of 8 µg ml<sup>-1</sup> polybrene followed by selection with appropriate antibiotic resistance markers.

**Cell proliferation assays.** MYC-ER HMECs were infected with pINDUCER-shRNA viruses at a multiplicity of infection (MOI) of 1.3–1.5, and transduced cells were seeded at a density of 3,000 onto 96-well black plates (Corning). MYC-ER HMECs with pINDUCER-shBUD31-3'UTR were treated with 300 nM 4-hydroxytamoxifen (4-OHT) to induce MYC hyperactivation, and with 32 ng ml<sup>-1</sup> dox (Sigma) to induce shBUD31 expression. SUM159 and MDA-MB-231-LM2 cells were infected with pINDUCER-shBUD31 virus (targeting the 3' UTR and coding region, respectively) at an MOI of 1.5, and seeded at a density of 1,000 and 2,000, respectively. Expression of shBUD31 in LM2 and SUM159 cells was induced with 1 µg ml<sup>-1</sup> dox. HMECs and breast cancer cells were re-fed every 3–4 days until cells reached confluence. At confluence, cells were fixed in 4% paraformaldehyde, and nuclei were stained with Hoechst3321 (1:1,000, Life Technologies). Nuclei were imaged and counted using the Celigo Imaging Cell Cytometer (Brooks).

For clonogenic assays, breast cancer or immortalized epithelial cells were seeded at low density (between 500 and 2,000 cells per plate, depending on the cell line) into 6-cm plates, four replicates per treatment group. MYC-ER HMECs with pINDUCER-shBUD31-3'UTR were treated with 8 ng ml<sup>-1</sup> dox and 300 nM 4-OHT, and MYC-ER HMECs treated with 10 or 20 nM SD6 were also cultured with 200 nM 4-OHT. Cells were re-fed every 4 days until colonies were macroscopic. The colonies were stained using Coomassie brilliant blue. Macroscopic colonies were quantified and normalized to vehicle-treated cells for each cell line.

For the P493-6 cell line with pmc-tet construct<sup>32</sup>, MYC was reduced by treating cells with 0.1 µg ml<sup>-1</sup> tetracycline (Sigma) for 72 h. MYC was induced by washing P493-6 cells with PBS twice, then culturing cells in RPMI-1640 medium with 10% Tet System Approved FBS (Clontech) and 1% GlutaMAX. P493-6 cells were treated with or without 100 nM SD6 and with or without 0.1 µg ml<sup>-1</sup> tetracycline for 4 days.

**Immunoprecipitation and mass spectrometry.** HeLa cells transduced with lentivirus encoding *BUD31* cDNA and non-transduced HeLa cells were collected, and nuclear extracts as well as whole-cell lysates were collected as described previously<sup>33</sup>. Lysates were treated with RNase A (500 µg ml<sup>-1</sup>) for 1 h on ice. For immunoprecipitations, nuclear and whole-cell extracts were ultracentrifuged at 100,000g, and incubated with 25 µg M2 Flag antibody (Sigma) for 1 h, followed by ultracentrifugation and incubation with Sepharose-CL4B Protein A beads (GE Healthcare). Beads were washed with NTN (50 mM Tris-Cl, pH 8.0, 150 mM NaCl and 0.5% NP-40), and immunocomplexes were resuspended in 1× Laemmli buffer and resolved on pre-cast 4–20% Novex Tris-Glycine gels (Life Technologies). Gels were minimally stained with Coomassie brilliant blue, cut into 8 molecular mass ranges, and digested with trypsin. Immunocomplexes were identified on a Thermo Fisher LTQ mass spectrometer, and data processing was performed as previously described<sup>33</sup>.

**Enrichment analysis.** Human GO annotation file (gene\_association.goa\_human.gz) was downloaded from <http://geneontology.org/GO.downloads.annotations.shtml> containing a GOC Validation date of 2 September 2013. Enrichment analysis was performed to consider the content of (1) *BUD31*-associated proteins, or

(2) genes with enhanced IR. Gene symbols annotated to *BUD31*-associated proteins were cross tabulated against all Gene Ontology annotations. Genes with enhanced IR were cross tabulated against the subset of Gene Ontology annotations for genes considered in this analysis. We used Fisher's exact test to determine *P* values for the proportion of genes overlapping each annotation set.

**BiFC.** *BUD31* was cloned into the pQCXIN-N-YFP fusion vector, in which the *BUD31* N terminus was fused to the N-terminal domain (residues 1–155) of Venus yellow fluorescent protein (YFP). Human splicing factor cDNAs were individually recombined into retroviral vectors with C-terminal Venus YFP (residues 156–239) tags at the N-terminal ends. SUM159 breast cancer cells were transduced with these bait and prey BiFC retroviruses, and cellular fluorescence was analysed by flow cytometry in triplicate.

**BUD31 mutagenesis.** Wild-type and mutant *BUD31* cDNAs were generated by gene synthesis (IDT DNA) and recombined into the pQCXIN-N-YFP fusion vector. Mutant *BUD31* consisted of substituting human *BUD31* amino acid residues 105–114 with an equivalent number of glycine residues (codon GGA).

**In vitro competition assay.** MYC-dependent SUM159 breast cancer cells with pINDUCER-shBUD31-3'UTR were transduced with viruses containing wild-type or mutant *BUD31* or negative control cDNA recombined into pQCXIN-N-YFP vectors. Infected, GFP<sup>+</sup> cells are mixed at an 80:20 ratio with non-transduced, GFP<sup>-</sup> parental cells and seeded into 96-well plates and treated either with or without dox (1 µg ml<sup>-1</sup>). At confluence, cells were passaged 1:10 and processed for flow cytometry. The *in vitro* competition assay was continued for two passages.

**Immunoblotting.** Cells were lysed in 1× SDS sample buffer (62.5 mM Tris-HCl, pH 6.8, 10% glycerol, 2% SDS, 2.5% β-mercaptoethanol) and heated at 95 °C for 12 min. The following antibodies were used for western blotting: Flag (Sigma, A8592), *BUD31* (ProteinTech, 11798-1-AP), *SF3B1* (Bethyl, A300-996A), *Prp19* (Bethyl, A300-101A), *U2AF1* (Bethyl, A302-079A), *SF3A1* (Bethyl, A301-603A), *EFTUD2* (Bethyl, A300-957A), *SNRPF* (Abcam, 154870), *HER2* (Millipore, 06-562), *EGFR* (Cell Signaling, 2232), cleaved caspase-3 (Cell Signaling, 9664), *RPS8* (Assay Biotechnology, R12-3466), *EIF2S1* (Abgent, AP13469s), *eIF3I* (p36) (Biolegend, 646701) and c-Myc (D84C12) (Cell Signaling, 5605). Vinculin (Sigma, V9131) and Ran (BD Biosciences, 610340) were used as loading controls.

**In vitro transcription.** Uniformly <sup>32</sup>P-UTP radiolabelled MINX pre-mRNA was *in vitro* transcribed from a BamHI-digested plasmid<sup>34</sup>, DNaseI (Ambion) treated and gel-isolated on a 8 M urea 6% polyacrylamide gel.

**In vitro splicing.** HeLa nuclear extracts used for *in vitro* splicing assays were made as described previously<sup>35</sup> from HeLa cells transduced with an inducible *BUD31*-targeting shRNA and grown in the presence or absence of 1 µg ml<sup>-1</sup> dox. Splicing reactions of 15 µl contained: 8 nM RNA substrate, 0.8 mM DTT, 1.7 mM magnesium acetate, 1.7 mM ATP, 17 mM phospho-creatine, 20 mM glycine, 1 U µl<sup>-1</sup> RNasin Plus (Promega), 3.7% PVA and 50 µg of HeLa nuclear extracts. Splicing reactions were incubated for indicated time points at 30 °C and stopped by digestion with proteinase K (Ambion) for 30 min at 45 °C followed by RNA purification. RNA purified from splicing reactions was electrophoresed on 8 M urea 8% polyacrylamide gels, then exposed to a phosphorimager screen (Typhoon Trio phosphorimager, GE Healthcare). Alternatively, RNA purified from *in vitro* splicing reactions was added to RT-PCR reactions as previously described<sup>36</sup> with primers in exons 1 and 2 of MINX (forward: 5'-CGGAATTCGAGCTCGCCC-3' and reverse: 5'-GGATCCCCACTGGAAAGA-3'). PCR products were run on 6% non-denaturing polyacrylamide gels and visualized after staining with ethidium bromide.

**Spliceosome complex formation assay.** *In vitro* splicing reactions were carried out as described above, placed on ice, and heparin was added to a final concentration of 2 µg µl<sup>-1</sup>. Reactions were incubated in the presence of heparin at 30 °C for 5 min and immediately loaded onto 0.75-mm non-denaturing 4% acrylamide–0.4% agarose composite gels. Gels were run at 250 V at room temperature in 1× tris-glycine running buffer for 3 h, then placed on Whatman paper and exposed to a phosphorimager cassette.

**RNA isolation and qRT-PCR.** RNA isolation was performed with the RNeasy Mini kit (Qiagen). Reverse transcription was performed using the High Capacity RNA-to-cDNA Master Mix (Applied Biosystems), and qPCR was performed using SYBR Green Master Mix (Applied Biosystems). The following primers were used: *BUD31* forward: 5'-ACCAACTTCGGGACGAACTG-3', reverse: 5'-CGG CCACTTCCAGCTT-3'; *EFTUD2* forward: 5'-CCTTCGTGTTGTCAGAGA GTGTCT-3', reverse: 5'-TGGGTTGGAGGTTGGTGAGT-3'; *SF3B1* forward: 5'-GTGGACAAATGGCGAAGAT-3', reverse: 5'-GAGCTTCATCAAGAGCT GCC-3'; *SNRPF* forward: 5'-GGGAATGGAGTACAAGGGCT-3', reverse: 5'-CC CAGATGTCAGACAAAGC-3'; *U2AF1* forward: 5'-ACGTTTAGCCAGACCA TTGC-3', reverse: 5'-TGTTCTGCATCTCCACATC-3'; *GAPDH* forward: 5'-CCTCCCGCTTCGCTCTCT-3', reverse: 5'-TGCGACGCAAAAGAAGAT-3'.

**RNA-seq.** pINDUCER11-shBUD31-3'UTR-infected MYC-ER HMECs were cultured for 72 h with/without 16 ng ml<sup>-1</sup> dox, and for 48 h with/without 300 nM



tamoxifen in triplicates. Total RNA was isolated using the RNeasy kit (Qiagen). RNA samples were rRNA depleted, and NGS libraries were constructed and sequenced as 75 bp paired-end reads by Illumina HiSeq 2000.

**Quality assessment of RNA-seq.** RNA-seq NGS reads quality was evaluated using FastQC application (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>).

**Alignment of RNA-seq data.** RNA-seq NGS reads were mapped using STAR RNASeq aligner (version 2.3.1). To improve mapping accuracy, the database file of splice junctions (<http://it-collab01.cshl.edu/shares/gingeraslab/www-data/dobin/STAR/STARgenomes/GENCODE/Old/genencode.v14.annotation.gtf.sjdb>) was supplied at the genome index generation step with command line option `-sjdbOverhang 7`, together with [http://it-collab01.cshl.edu/shares/gingeraslab/www-data/dobin/STAR/STARgenomes/GENCODE/Old/hg19\\_Gencode14.overhang75/](http://it-collab01.cshl.edu/shares/gingeraslab/www-data/dobin/STAR/STARgenomes/GENCODE/Old/hg19_Gencode14.overhang75/) and default parameters. Duplicate reads were marked with the MarkDuplicates function of the Picard-tools software package (<http://picard.sourceforge.net;version 1.107>) using default settings.

**Intron–exon junction definition.** To prevent confounding effects in our analysis of IR within HMECs, we confined our analyses to exons in non-overlapping genes that are included within all isoforms of a given gene (75,623 junctions in 6,861 genes).

Intron–exon junctions were obtained using the University of California Santa Cruz Genome Browser 'knownGene' table (downloaded 4 June 2014). Constitutive junctions were defined as junctions that (1) appear in each transcript annotated to a given gene symbol, (2) do not overlap with any transcript annotated to a different gene symbol, and (3) do not mark the start or stop of a transcript.

**Junction IR calculation.** Because analysis of intronic reads may be influenced by the presence of stable RNAs within introns and/or spliced lariats, we calculated junction IR as the ratio of exon–intron reads to exon–exon reads, restricting the analysis to reads directly spanning exon–intron or exon–exon junction sequences. We used R together with the Rsamtools package to calculate IR. In brief, for each intron–exon junction, we extracted all non-duplicate reads overlapping this junction. Next, we assigned these reads into two categories: (1) 'intronic' if the read mapped to at least the first base of the intron, (2) 'exonic' if none of the bases of the read mapped to the first base of the intron and at least one base mapped to a subsequent exon. We counted the total number of reads assigned into each category for each junction. IR was calculated as:

$$IR_{j,i} = \log_2 \frac{I_{j,i} + 1}{E_{j,i} + 1}$$

in which  $IR_{j,i}$  represents the IR score for junction  $j$  in sample  $i$ , and  $I_{j,i}$  and  $E_{j,i}$  refer to the count of reads classified as intronic and exonic for junction  $j$  in sample  $i$ , respectively. To avoid ratios with 0 in the denominator, we added 1 to each of these counts. The scripts used to conduct this calculation are available on request. We restricted all following analyses to intron–exon junctions with an average of at least 25 total (intronic and exonic) reads in the control and MYC-hyperactivated samples.

**Gene IR calculation.** For the cumulative distribution analyses, the mean IR score for all junctions in a gene were averaged.

**Gene annotation.** A custom gene annotation file was generated to correspond to the set of intron–exon junctions considered in the IR analysis. In brief, exons were defined as: (1) an exon flanked by two junctions annotated to the same symbol, and (2) an exonic region flanked by one junction and conserved across all transcripts annotated to the same symbol.

**Statistical analysis of RNA-seq.** Statistical analyses were performed using the open source statistical programming environment 'R'. Empirical cumulative distributions of IR scores were compared using two-sided Kolmogorov–Smirnov test and Wilcoxon test.

**Permutation-based test of significance.** The significance of the difference of empirical cumulative distributions of junction-level IR scores was evaluated using a permutation-based approach. The null hypothesis was that splicing perturbations had no effect on IR changes in the MYC-normal and MYC-hyperactivated states. To model this null hypothesis, the treatment information was blinded to the assignments of MYC activity. We generated a third control sample by randomly selecting half of the junctions from samples 1-control and 2-control. Control and MYC samples were grouped as 6 'normal' samples, and LowBUD31 and Myc\_LowBUD31 samples were grouped as 'splicing perturbed' samples. Next, we comprehensively generated all possible normal and splicing perturbed contrasts by subtracting the average of each junction IR score of three normal samples from that of the splicing perturbed group. The empirical distribution of all possible double differences was generated and used to assign significance to the original observations. An analogous approach was used to evaluate the difference of empirical cumulative distributions of gene-level IR scores.

**qPCR IR validation assay.** Amplification reactions were prepared using SYBR Select Master Mix (Applied Biosystems) according to the manufacturer's instructions with final primer concentration of 300 nM. Reactions were performed using a StepOnePlus Real Time PCR System (Applied Biosystems) with an initial incubation at 95 °C for 10 min followed by 40 cycles of 15 s at 95 °C and 1 min at 60 °C. Primers were designed using Primer3 (available at <http://bioinfo.ut.ee/primer3/>) and assessed for quality using Beacon Designer (<http://www.premier-bio.com/qpcr/>) and UNAFold (<https://www.itdtna.com/UNAFold>). Primers used for each reaction were: *HTRA1*\_IE forward: 5'-GCGTTCATTTTAAGGTGCTACAGG-3', reverse: 5'-TGGGCATTGTGTCACGATCAGT-3'; *HTRA1*\_EE forward: 5'-GACGTGGTGGAGAAGATCGC-3', reverse: 5'-AAACCCAGACCCTAGCCCA-3'; *PRPF19*\_IE forward: 5'-TCCCCTTGTTGACCTTCTCT-3', reverse: 5'-AAGATCTCCGTCCATTGTTTGC-3'; *PRPF19*\_EE forward: 5'-AGAACTTTAAGACTTTGCAGCTGG-3', reverse: 5'-TCTCCGTCCATTGTTTG CAGA-3'; *UBALD2*\_IE forward: 5'-GCTGCGTTTCCGTACTCCG-3', reverse: 5'-GTGGTGGCTGTTGGGAATGT-3'; *UBALD2*\_EE forward: 5'-CAGTTGCTGCAGGCGGCC-3', reverse: 5'-TGGAGAAGCTGCTCAGCGC-3'.

Ultramer oligonucleotides (Integrated DNA Technologies) were synthesized to match the predicted amplicon of each primer pair, and standard curves were generated for each reaction using threefold serial dilutions of these control templates ranging in concentration from  $4.0 \times 10^{-12}$  M to  $1.6 \times 10^{-14}$  M. The sequence AAGAA was added to both the 5' and 3' ends of each template to facilitate primer binding. Control template sequences for intron–exon (IE) and exon–exon (EE) regions were as follows: *HTRA1*\_IE: 5'-AAGAAGCGTTCATTTAAGGTGCTACAGGCTTAAGTGTGTAATCCTTTGGATTTAGGCTTCGTTTTCTAAACGAGAGGTGCCGCTAGTGGGTCTGGGTTTATTGTGTCGGAAGATGGACTGATCGTGACAAATGCCCAAGAA-3'; *HTRA1*\_EE: 5'-AAGAAGACGTGGTGGAGAAGATCGCCCTGCCGTGGTTCATATCGAATTGTTTCGCAAGCTTCCGTTTTCTAAACGAGAGGTGCCGCTG-GCTAGTGGGTCTGGGTTTAAGAA-3'; *PRPF19*\_IE: 5'-AAGATCCCCTTGTGTGACCTTCTCTCTTTCTATTTCTGGCAGGTAAAGTCACTGATCTTTGACCAGAGTGGTACCTACCTGGCTCTTGGGGGCACGGATGTCCAGATCTACATCTGCAACAAATGGACGGAGA-AAGAA-3'; *UBALD2*\_IE: 5'-AAGAAGTTCGCTTCTGACTCCGCTGCCCGCCGCTGCTACTGCCCTGTTTGTCCGACACCGCGCTGAGCAGTCTTCCAAAGAAAC-CAACA TTCCCAACAGCCACCACAAGAA-3'; *UBALD2*\_EE: 5'-AAGAACAGTTGCTGCAGGCGGCCCACTGGCAGTTCGAGACCGCGCTGAGCAGGTTCTTCCAAA-GAA.

$C_t$  values from each reaction were interpolated on the standard curve generated using the corresponding control template to approximate the concentration of cDNA template in each experimental sample. These values were then reported as the ratio of intron–exon to total (IE + EE) cDNA template in each sample.

**Transcription pulse assay.** MYC-ER HMECs with pINDUCER11-shBUD31-3'UTR were cultured with/without 16 ng ml<sup>-1</sup> dox and/or with/without 300 nM 4-OHT. Cells were pulsed with 500 µM 4-thiouridine (4-SU, Sigma) for 2 h, and collected for total RNA using RNeasy mini kit (Qiagen). 4-SU-labelled RNA was purified from 20 µg total RNA. Isolation of newly transcribed RNA was performed as described<sup>37</sup> using 100 µl streptavidin beads (Miltényi Biotec).

**Poly(A)<sup>+</sup> RNA isolation.** Dynabeads Oligo(dT)<sub>25</sub> (Life Technologies) were equilibrated with 50 µl lysis/binding buffer, and total RNA was heat denatured (70 °C for 2 min) before binding poly(A)<sup>+</sup> RNA to Dynabeads. Isolation of mRNA was performed according to manufacturer's instructions. Poly(A)<sup>+</sup> RNA concentrations were measured with a fluorescence plate reader (Molecular Devices) using Quant-iT RiboGreen reagent (Life Technologies).

**Poly(A)<sup>+</sup> RNA LNA FISH.** pINDUCER11-shBUD31-3'UTR-transduced MYC-ER HMECs were seeded onto collagen-coated 8-well glass chamber slides and cultured with/without 16 ng ml<sup>-1</sup> dox and with/without 300 nM tamoxifen. Cells were treated with/without 2 µg ml<sup>-1</sup> actinomycin D (Gibco) or DMSO for 5 h before fixation in 4% formaldehyde and 5% acetic acid in PBS for 15 min at room temperature. Fixed cells were washed with PBS, permeabilized with proteinase K (5 µg ml<sup>-1</sup>, Life Technologies) and treated with/without RNase A (100 µg ml<sup>-1</sup>, Sigma) for 30 min at 37 °C in PBS. Dehydration of the cells was performed with 70%, 95% and 100% ethanol solutions. FITC-labelled oligo(dT)<sub>25</sub> locked nucleic acid (LNA) probes were heated to 90 °C for 4 min, then cooled to hybridization temperature (55 °C). Dehydrated and dried cells were incubated in 40 nM of LNA probes in hybridization buffer (50% formamide, 2 × SSC, 50 mM NaPi, pH 7.0, 10% dextran sulphate) overnight at 55 °C. Chamber slides were washed with 5 × SSC, 1 × SSC, 0.2 × SSC and PBS, and dehydrated before counterstaining with DAPI and mounting with Fluoromount-G (Southern Biotech). Cells were imaged using a Nikon Ti-E inverted microscope with 40× air objective and Andor

Zyla 4.2 sCMOS camera. For each treatment condition and actinomycin time point,  $\geq 150$  cells were analysed for mean FITC intensity. Cellular FITC values were adjusted for background fluorescence by subtracting the mean extra-cellular pixel value. Image analysis was performed using Nikon Elements.

**Luminescent apoptosis assays.** Caspase-3/7 activity was assessed in MYC-ER HMECs and breast cancer cell lines by incubating Caspase-Glo 3/7 Reagent with cells in triplicate wells of a 96-well plate and measuring luminescence with a plate reader (Molecular Devices). Luminescence was normalized using cell numbers determined by Hoechst3321 staining of a duplicate plate, followed by nuclei counting using the Celigo Imaging Cell Cytometer (Brooks).

**Tumorigenicity and metastasis assays.** SUM159 breast cancer cells were transduced with pINDUCER11-shBUD31-3'UTR virus and analysed by flow cytometry to confirm  $>98\%$  transduction. In total  $8 \times 10^6$  transduced cells were injected with matrigel (BD Biosciences) subcutaneously into the flank of four-week-old female athymic nude Foxn1-nu mice (Harlan Labs). Tumour volume was measured using calipers, and once tumours achieved  $150 \text{ mm}^3$ , mice were randomized onto and maintained on sucrose water ( $-dox$ ) or sucrose water with dox ( $+dox$ ).

For mixed population experiments, MDA-MB-231-LM2 breast cancer cells were individually transduced with pINDUCER11-shRNAs targeting the indicated genes at an MOI appropriate to transduce all cells (1.3–1.5). The individual populations were mixed at equal ratios *in vitro* and expanded before injection. Around  $3 \times 10^6$  or  $2 \times 10^5$  mixed population cells were injected subcutaneously into the right flank or into the lateral tail vein of four-week-old female athymic nude Foxn1-nu mice (Harlan Labs), respectively. Subcutaneous tumour volume was measured with calipers over time. Mice were randomized onto sucrose water ( $-dox$ ) or sucrose water with dox ( $+dox$ ) after tumours exceeded  $150 \text{ mm}^3$ . Lung metastatic progression was monitored and quantified using noninvasive bioluminescence as described previously<sup>19</sup>. When tumours reached  $1,000 \text{ mm}^3$  or the total luminescence flux reached  $1 \times 10^9$ , genomic DNA from dissected tumours or lungs were collected using the QIAamp DNA mini kit (Qiagen). qPCR was performed with SYBR Green PCR Master Mix (Life Technologies) using manufacturer's recommendations and the following primers. Experimental target  $C_t$  values were normalized to the *TRE*  $C_t$  values, and NCOR2 was used as a negative control.

The following primers were used. *BUD31* forward: 5'-TGGAAGACATCTGCGTGGTATT-3', reverse: 5'-CGCGCAAACCTAAAGGCATA-3'; *SF3B1* forward: 5'-GCCGTATCATTAGTACGCCATA-3', reverse: 5'-TCGATCCTAGGACGGGTAT-3'; *MYC* forward: 5'-GCCGGCCATATTTTCACTTC-3', reverse: 5'-CACACCTACCGAAAAACAAAC-3'; *NCOR2* forward: 5'-AACTTCCGGTGTGTCGTTT, reverse: 5'-CGCGTCTAGGTAATACGACTCA-3'; *TRE* forward: 5'-TGTACGGTGGGAGGCCTATATAA, reverse: 5'-GCGTCTCCAGGCGATCTG-3'.

For SD6 drug infusion studies,  $3 \times 10^6$  or  $2 \times 10^5$  MDA-MB-231-LM2 breast cancer cells were injected into the flank or lateral tail vein of four-week-old female athymic nude Foxn1-nu mice (Harlan Labs), respectively. For mice with subcutaneous tumours, jugular vein catheters (SAI Infusion Technologies) were surgically implanted into each mouse 13–16 days after injection, and were randomized to receive vehicle ( $n = 11$ ) or SD6 ( $n = 10$ ) infusion. Tail-vein-injected mice were randomized to receive vehicle ( $n = 7$ ) or SD6 ( $n = 6$ ) infusion 1 day after tail vein injections. Animals received daily infusions of vehicle (10% 2-hydroxypropyl- $\beta$ -cyclodextrin dissolved in 50 mM  $\text{Na}_2\text{HPO}_4/\text{NaH}_2\text{PO}_4$ , pH 7.4) or 50 mg  $\text{kg}^{-1}$  of SD6 for 20 consecutive days (subcutaneous cohort) or 10 consecutive weekdays (tail vein cohort). Mice were infused via jugular catheter at a rate of  $3.5 \mu\text{L min}^{-1}$  with a Fusion 200 Touch Syringe Pump (SAI Infusion Technologies). The total volume infused did not exceed 500  $\mu\text{L}$  per day. Subcutaneous tumour volumes were monitored with calipers, and lung metastatic progression was monitored with noninvasive bioluminescence. Mice were euthanized once tumours reached  $2,000 \text{ mm}^3$  or the total luminescence flux reached  $1 \times 10^9$ . In progression-free survival analyses, progression is defined as fivefold increase in pulmonary bioluminescence relative to initial values or fourfold increase in subcutaneous tumour volume relative to its volume at time of randomization.

Investigators responsible for monitoring and measuring the xenografts of individual tumours were not blinded. Simple randomization was used to allocate animals to experimental groups. All animal studies were performed in accordance with institutional and national animal regulations. Animal protocols were approved by the Institutional Animal Care and Use Committee at Baylor College of Medicine.

Power analysis was used to determine appropriate sample size to detect significant changes in animal survival, which were based on previous survival analyses in our laboratory. All animals were included in analyses.

**Pooled shRNA screens in breast cancer cell lines.** Pooled shRNA screens were performed on 68 breast cancer lines and four non-malignant immortalized mammary epithelial lines, essentially as described<sup>38</sup>. In brief, cells are infected with

a lentiviral shRNA library at a MOI of 0.3, and passaged under standard conditions. At 4 and 8 doublings, respectively, DNA is isolated and hybridized to a customized chip to assess shRNA dropout. A detailed description of the results of these screens will be published separately (R.M., A. S. and B.G.N., manuscript in preparation).

**Correlation between MYC dependency and spliceosome dependency.** First, to calculate MYC dependency scores using probable on-target hairpins, we used assay observations associated with 3 hairpins incorporated into the first ATARIS<sup>39</sup> solution for MYC. MYC dependency scores were generated using a hierarchical linear model, with pooled shRNA screen observations as the independent variable and two regression covariates: initial signal intensity (with coefficient  $\beta_0$ ) and linear time-course dropout trend (with coefficient  $\beta_1$ ). The dropout trend is calculated for each cell line separately, resulting in a per-cell-line MYC dropout score (the value of coefficient  $\beta_1$ ).

Second, the MYC dependency score was used in a hierarchical linear model to search for associations with the essentiality of other genes (such as spliceosome encoding genes). This model uses pooled shRNA screen observations as the independent variable and three regression covariates: initial assay signal intensity (with coefficient  $\beta_1$ ), linear time-course dropout trend (with coefficient  $\beta_1$ ), and an interaction term between dropout trend and MYC dependency score (with coefficient  $\beta_M$ ). The *P*-value associated with the interaction term  $\beta_M$  is used to determine whether a significant association exists. A detailed description of this approach will be published elsewhere (A.S., R.M. and B.G.N., manuscript in preparation).

A summary statistic using results from the single-gene analyses was used to test the significance of the association between MYC dependency and the essentiality of a gene set. For a gene set containing genes  $g$ , we calculate the gene set summary statistic as

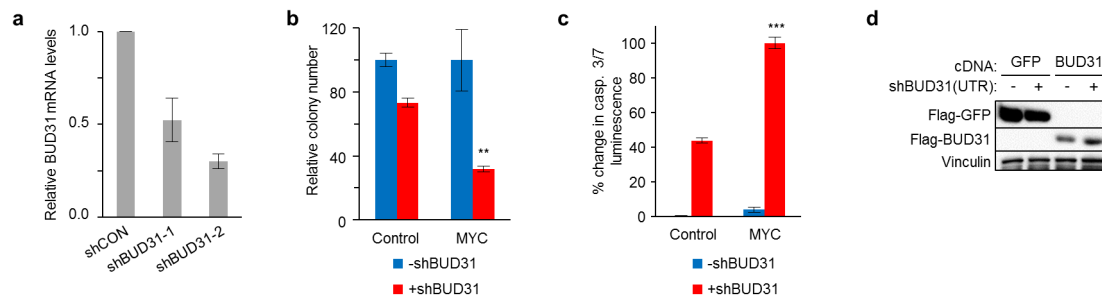
$$-\sum_{\text{gene } g} \text{sign}(\beta_M) \log_{10}(P \text{ value}(\beta_M))$$

in which  $\text{sign}(\beta_M)$  and  $P\text{value}(\beta_M)$  indicate the values associated with the regression coefficient  $\beta_M$ . The resulting metric, termed a siMEM (mixed-effect model) score, indicates the significance and correlation between sensitivity to MYC shRNAs and sensitivity to a group of shRNAs targeting a gene set (such as those targeting the spliceosome). A gene set (for example, spliceosome genes) for which a substantial number of genes are significantly associated with MYC dependency, and all with the same direction (sign) of association, will have a large positive score.

When calculated for the gene set consisting of the core spliceosome, this value summarizes the direction and strength of the significance observed across genes in the spliceosome. To determine whether this observation is significant, the same statistic is calculated for 100,000 randomly drawn gene sets of the same size as the core spliceosome, yielding the null distributions of gene set summary statistics in Fig. 4b, c.

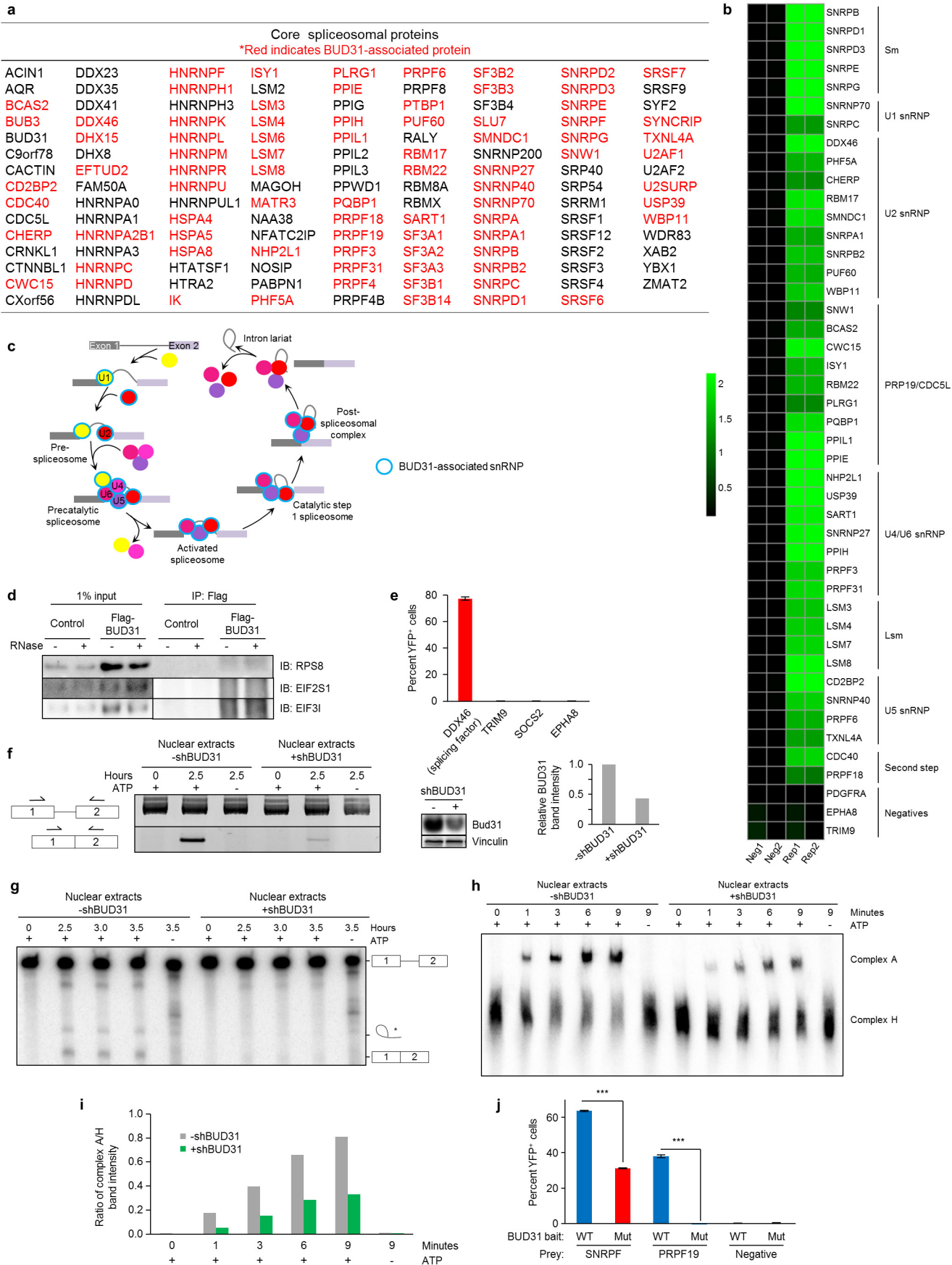
**Statistical analysis.** All experiments were performed on biological replicates unless otherwise specified. Sample size for each experimental group/condition is reported in the appropriate figure legends and methods. For cell culture experiments, sample size was not predetermined, and all samples were included in analyses. For significance testing, analyses were chosen if data met the assumptions of the tests. Data was checked for comparable variance before statistical analysis. Statistically significant differences between control and experimental groups were determined using two-tailed unpaired Student's *t*-test, one-way ANOVA with Tukey–Kramer minimum significant difference test, Mann–Whitney test, Kolmogorov–Smirnov test, Wilcoxon test, permutation-based test of significance, and log-rank test as indicated in the appropriate figure legend and methods text.

- Meerbrey, K. L. *et al.* The pINDUCER lentiviral toolkit for inducible RNA interference *in vitro* and *in vivo*. *Proc. Natl Acad. Sci. USA* **108**, 3665–3670 (2011).
- Schuhmacher, M. *et al.* Control of cell growth by c-Myc in the absence of cell division. *Curr. Biol.* **9**, 1255–1258 (1999).
- Malovannaya, A. *et al.* Streamlined analysis schema for high-throughput identification of endogenous protein complexes. *Proc. Natl Acad. Sci. USA* **107**, 2431–2436 (2010).
- Zapp, M. L. & Berger, S. M. Evidence for nuclear factors involved in recognition of 5' splice sites. *Nucleic Acids Res.* **17**, 2655–2674 (1989).
- Dignam, J. D., Lebovitz, R. M. & Roeder, R. G. Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. *Nucleic Acids Res.* **11**, 1475–1489 (1983).
- Echeverria, G. V. & Cooper, T. A. Muscleblind-like 1 activates insulin receptor exon 11 inclusion by enhancing U2AF65 binding and splicing of the upstream intron. *Nucleic Acids Res.* **42**, 1893–1903 (2014).
- Dölken, L. *et al.* High-resolution gene expression profiling for simultaneous kinetic parameter analysis of RNA synthesis and decay. *RNA* **14**, 1959–1972 (2008).
- Marcotte, R. *et al.* Essential gene profiles in breast, pancreatic, and ovarian cancer cells. *Cancer Discov.* **2**, 172–189 (2012).
- Shao, D. D. *et al.* ATARIS: computational quantification of gene suppression phenotypes from multisample RNAi screens. *Genome Res.* **23**, 665–678 (2013).



**Extended Data Figure 1 | Validation of *BUD31* as a MYC-synthetic lethal gene in HMECs.** **a**, qRT-PCR analysis of *BUD31* mRNA level (mean  $\pm$  s.d.,  $n = 3$  biological replicates). **b**, Clonogenicity of MYC-ER HMECs with or without MYC hyperactivation or *BUD31* depletion (mean  $\pm$  s.e.m.,  $n = 4$

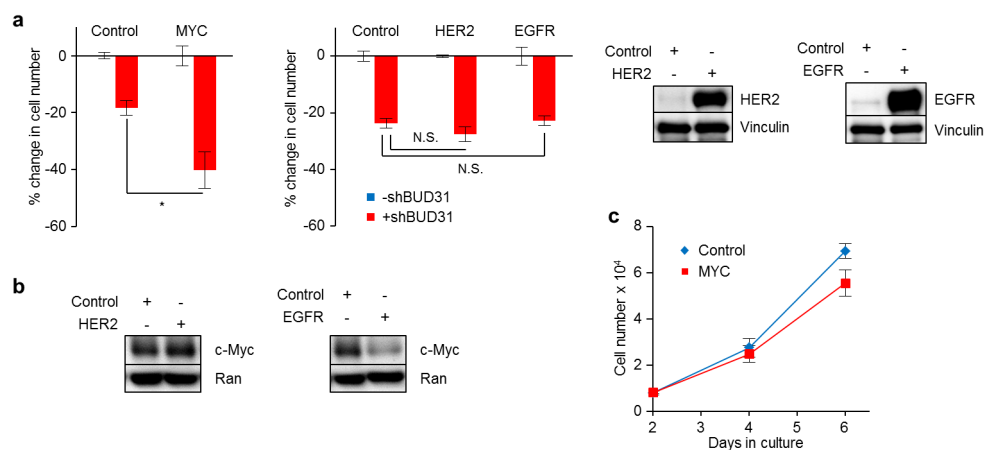
biological replicates,  $**P < 0.01$ , two-tailed Student's *t*-test). **c**, Caspase-3/7 activation by caspase luminescence assay (mean  $\pm$  s.e.m.,  $n = 3$ ,  $***P < 0.001$ , one-way ANOVA). **d**, Flag-tagged protein levels in MYC-ER HMECs in which vinculin was used as a loading control.





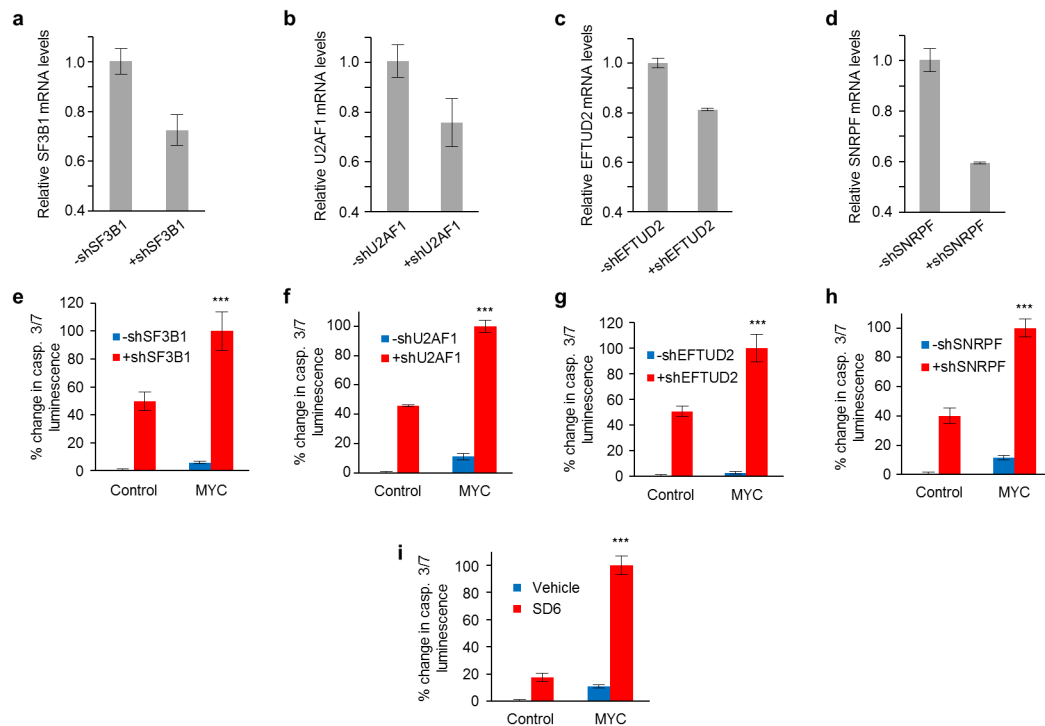
**Extended Data Figure 2 | BUD31 interacts with core spliceosomal factors and is required for spliceosomal assembly and pre-mRNA splicing.** **a**, 134 core spliceosomal proteins are listed. Proteins in red are shown to interact with BUD31, as discovered by Flag-BUD31 immunoprecipitation mass spectrometry and BUD31 BiFC. **b**, Heat map of BUD31-interacting spliceosomal proteins, organized by spliceosome sub-complexes. A black-green colour scale depicts normalized BiFC interaction values between spliceosomal proteins and negative control protein (technical replicates in two left lanes) and BUD31 (technical replicates in two right lanes). **c**, Spliceosomal snRNPs (coloured circles) interact in a stepwise manner to excise intronic sequences from pre-mRNA. snRNPs with proteins identified from the BUD31 immunoprecipitation and mass spectrometry are noted (blue outline) to be BUD31-associated. **d**, Co-immunoprecipitation of Flag-BUD31 for non-spliceosomal proteins. Input and immunoprecipitation blots probed by EIF2S1 and EIF3I were taken at different exposures to minimize background signal. **e**, Interaction between N-YFP-tagged BUD31 and C-YFP-tagged spliceosomal (DDX46) or cytoplasmic proteins (TRIM9, SOCS2 and EPHA8) was assessed by cellular

fluorescence (mean  $\pm$  s.e.m.,  $n = 3$  technical replicates). **f**, Nuclear extracts with or without BUD31 knockdown were incubated with pre-mRNA substrate, and RT-PCR of unspliced RNA (top) and spliced RNA (bottom) was performed, using primers at the indicated arrows (left). BUD31 protein levels in the nuclear extracts were normalized to vinculin expression (middle) and quantified (right). **g**, Radioactively labelled pre-mRNA (MINX) was incubated with nuclear extracts with or without BUD31 depletion. RNA purified from the splicing reaction was run on a denaturing gel and imaged by autoradiography. The identities of prominent bands are based on size. Asterisk denotes putative intron-lariat band. **h**, After *in vitro* splicing was performed as described previously, products were electrophoresed on native gel, and spliceosome complexes were visualized by autoradiography. Complex A and nonspecific H complexes are labelled. **i**, Phosphorimager quantification of the ratio of RNA in complex A compared to that in complex H. **j**, Interaction between N-YFP-tagged wild-type (WT) or mutant BUD31 and C-YFP-tagged splicing factors was assessed by cellular fluorescence (mean  $\pm$  s.e.m.,  $n = 2$  technical replicates, \*\*\* $P < 0.001$ , two-tailed Student's *t*-test).



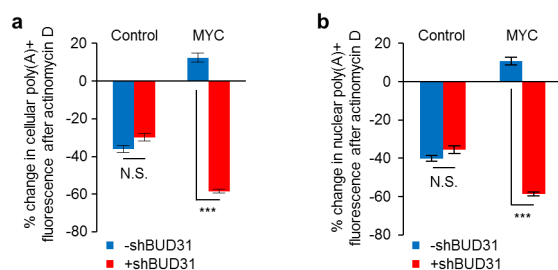
**Extended Data Figure 3 | HMECs with oncogenic activation of HER2 and EGFR do not require BUD31.** **a**, Cell number changes in HMECs with inducible shBUD31 and constitutive HER2 or EGFR expression (mean  $\pm$  s.e.m.;  $n = 4$  technical replicates;  $*P < 0.05$ , two-tailed Student's

*t*-test). HER2 and EGFR protein is normalized to vinculin (right). **b**, MYC protein levels in HMECs with constitutive HER2 or EGFR expression. **c**, MYC induction by tamoxifen in MYC-ER HMECs does not increase cell proliferation over time (mean  $\pm$  s.e.m.,  $n = 8$  technical replicates).



**Extended Data Figure 4 | Partial knockdown of core splicing factors is MYC-synthetic lethal in HMECs.** **a–d**, mRNA levels for core splicing factors SF3B1 (**a**), U2AF1 (**b**), EFTUD2 (**c**) and SNRPF (**d**) were evaluated by qRT-PCR (mean  $\pm$  s.d.,  $n = 3$  technical replicates). **e–i**, Caspase-3/7

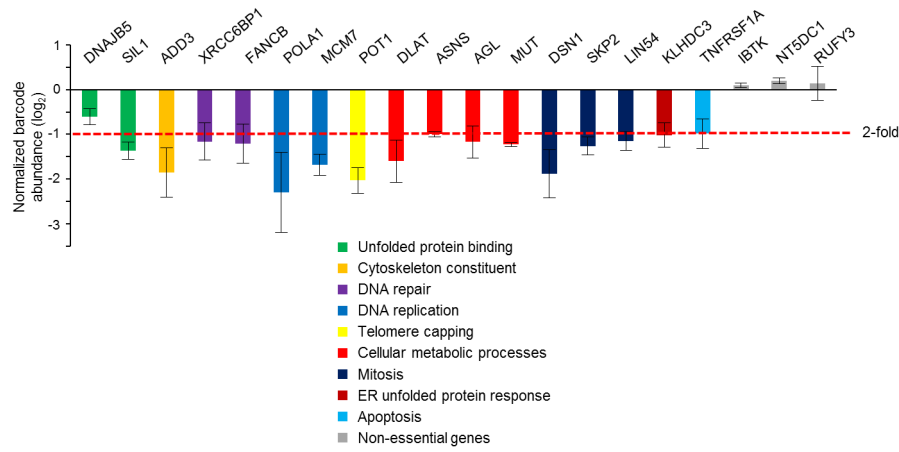
luminescence in MYC-ER HMECs with partial suppression of core spliceosomal proteins (**e–h**) or spliceosome inhibitor SD6 (**i**) (mean  $\pm$  s.e.m.,  $n = 3$  technical replicates, \*\*\* $P < 0.001$ , one-way ANOVA).



#### Extended Data Figure 5 | BUD31 loss in MYC-hyperactivated cells

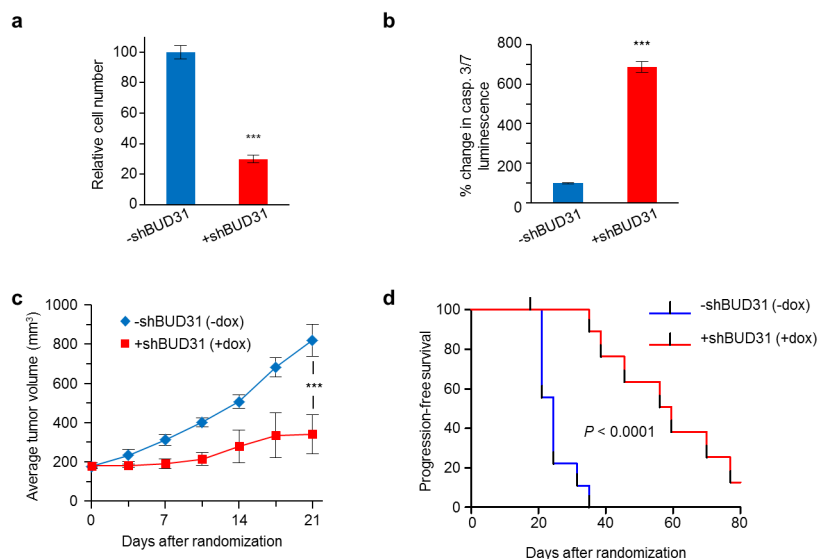
**destabilizes mRNA.** **a, b,** MYC-ER HMECs with inducible shBUD31 treated with actinomycin D for 5 h were labelled with oligo(dT)<sub>25</sub> LNA probes via fluorescence *in situ* hybridization. Cellular FITC intensity was assessed within cellular (**a**) and nuclear (DAPI+) (**b**) regions. Data are represented as the difference in cellular FITC intensity between 0 and 5 h of actinomycin D treatment in each cell state (mean  $\pm$  s.e.m.,  $n = 150$ , \*\*\* $P < 0.001$ , two-tailed Student's  $t$ -test).





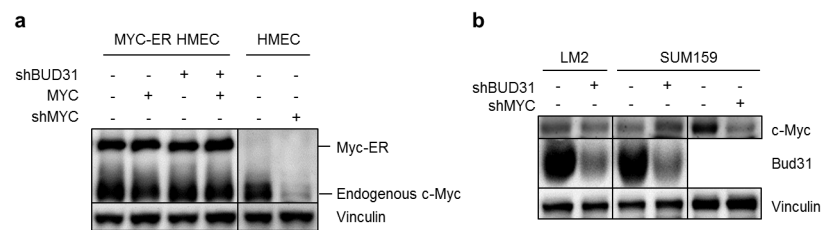
**Extended Data Figure 6 | BUD31 depletion in MYC-hyperactivated cells enhances intron retention and decreases expression of cell-essential genes.** In MYC-hyperactive cells, 17 representative genes display increased IR and decreased steady-state RNA levels after BUD31 knockdown. Depletion of these

genes by shRNA decreased cell viability (mean barcode abundance  $\pm$  s.e.m.). Twofold decrease in barcode abundance is noted by the dashed red line. All values are reflective of three biological replicates, and genes are colour-coded based on their Gene Ontology term annotation.



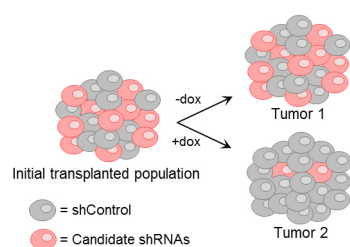
**Extended Data Figure 7 | MYC-dependent breast cancer cells require BUD31 for *in vitro* and *in vivo* growth.** **a**, Relative cell number of SUM159 cells with doxycycline-inducible shBUD31 *in vitro* (mean  $\pm$  s.e.m.,  $n = 8$  technical replicates, \*\*\* $P < 0.001$ , two-tailed Student's  $t$ -test). **b**, Caspase-3/7 luminescence in BUD31-depleted SUM159 cells (mean  $\pm$  s.e.m.,  $n = 3$  technical replicates, \*\*\* $P < 0.001$ , two-tailed Student's  $t$ -test). **c**, **d**, SUM159

cells engineered with dox-inducible shBUD31 were subcutaneously transplanted into mice and randomized onto dox treatment ( $-dox$   $n = 10$ ,  $+dox$   $n = 9$ ). Loss of BUD31 in SUM159 xenografts inhibits tumour growth (mean  $\pm$  s.e.m., \*\*\* $P < 0.001$  at day 21, two-tailed Student's  $t$ -test) (**c**) and prolongs progression-free survival (**d**) in nude mice ( $P$ -value, log-rank test).



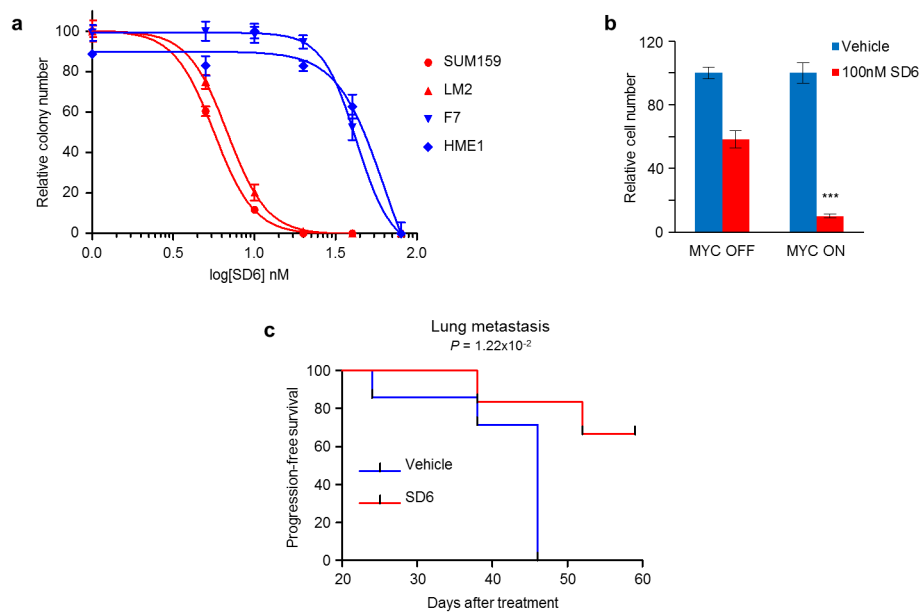
**Extended Data Figure 8 | BUD31 depletion does not affect levels of MYC protein.** **a**, MYC protein levels in MYC-ER HMECs with inducible shBUD31 expression normalized to vinculin expression. To confirm specificity of MYC antibody, HMECs without the MYC-ER construct were engineered

to express inducible MYC shRNA. **b**, MYC protein levels in SUM159 and LM2 cells with inducible shBUD31 normalized to vinculin expression. To confirm specificity of MYC antibody, SUM159 cells were engineered to express inducible MYC shRNA.



**Extended Data Figure 9 | Schematic for *in vivo* barcode-based competition assay.** LM2 cells transduced with inducible shRNAs targeting negative control genes or candidate genes were mixed at an equal ratio. This mixed population was transplanted into mice, and tumours were allowed to form in the presence or absence of dox. At the experimental endpoint, genomic DNA was isolated for comparisons of relative barcode (shRNA) abundance in tumour genomic DNA.





**Extended Data Figure 10 | Spliceosome inhibitor SD6 inhibits MYC-dependent cancer cells *in vitro* and *in vivo*.** **a**, MYC-dependent breast cancer cells (SUM159 and LM2) and MYC-normal immortalized epithelial cells (F7 and HME1) were cultured with SD6 at low density and analysed for clonogenic growth. **b**, MYC-repressible human B-cell line P493-6 was treated with or without 100 nM SD6 in the absence or presence of MYC hyperactivation for

four days, and cells were counted for relative cell number changes (mean  $\pm$  s.e.m.,  $n = 3$  biological replicates, \*\*\* $P < 0.001$ , one-way ANOVA). **c**, Kaplan–Meier survival analysis of nude mice with pulmonary seeding of LM2 cells treated with or without SD6 for 10 days (vehicle  $n = 7$ , SD6  $n = 6$ ,  $P$ -value by log-rank test).

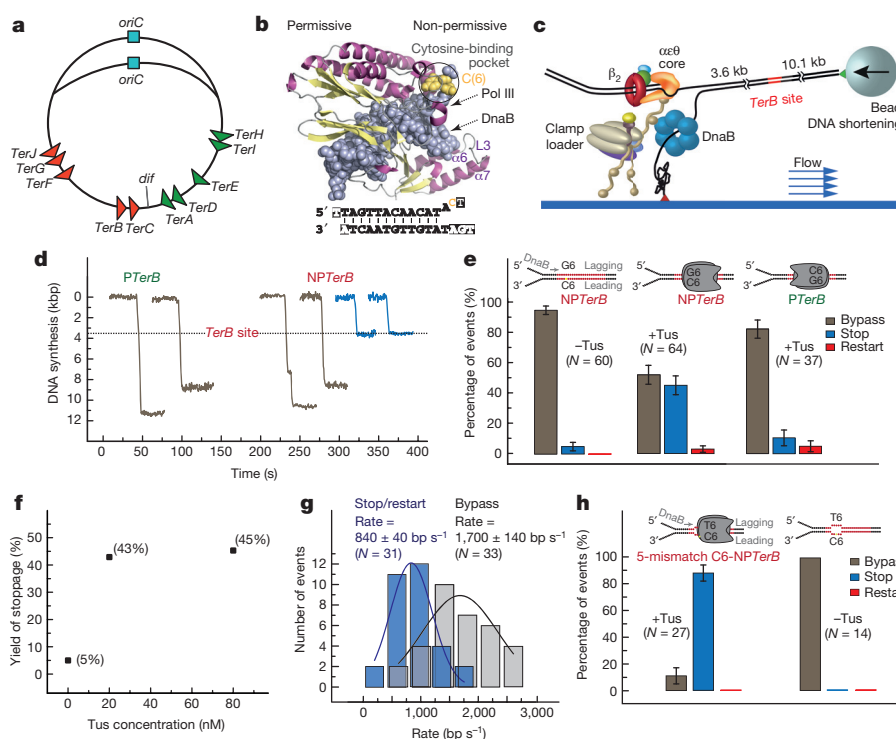
# Replisome speed determines the efficiency of the Tus–Ter replication termination barrier

Mohamed M. Elshenawy<sup>1</sup>, Slobodan Jergic<sup>2</sup>, Zhi-Qiang Xu<sup>2</sup>, Mohamed A. Sobhy<sup>1</sup>, Masateru Takahashi<sup>1</sup>, Aaron J. Oakley<sup>2</sup>, Nicholas E. Dixon<sup>2</sup> & Samir M. Hamdan<sup>1</sup>

In all domains of life, DNA synthesis occurs bidirectionally from replication origins. Despite variable rates of replication fork progression, fork convergence often occurs at specific sites<sup>1</sup>. *Escherichia coli* sets a ‘replication fork trap’ that allows the first arriving fork to enter but not to leave the terminus region<sup>2–5</sup>. The trap is set by oppositely oriented Tus-bound Ter sites that block forks on approach from only one direction<sup>3–7</sup>. However, the efficiency of fork blockage by Tus–Ter does not exceed 50% *in vivo* despite its apparent ability to almost permanently arrest replication forks *in vitro*<sup>8,9</sup>. Here we use data from single-molecule DNA replication assays and

structural studies to show that both polarity and fork-arrest efficiency are determined by a competition between rates of Tus displacement and rearrangement of Tus–Ter interactions that leads to blockage of slower moving replisomes by two distinct mechanisms. To our knowledge this is the first example where intrinsic differences in rates of individual replisomes have different biological outcomes.

In the circular *E. coli* chromosome, two replication forks move from the replication origin to converge opposite in a region that contains ten 23-base-pair Ter (termination) sites and the dif site for chromosome segregation<sup>3–7</sup> (Fig. 1a). The Ter sites are arranged in two oppositely



**Figure 1 | Fate of the *E. coli* replisome upon encountering Tus–TerB.**

**a**, Polarity of the replication fork trap. Replication occurs bidirectionally from *oriC*. Each fork passes through the first five permissive (P) Ter sites, but is arrested on encounter with one of the next five non-permissive (NP) sites. **b**, Structure of the Tus–Ter locked complex (PDB: 2I06)<sup>6</sup>. Strand separation of C6 (yellow) at the NP face induces its flipping into a specific binding pocket on Tus. **c**, Schematic of the single-molecule setup for observing leading-strand synthesis, which converts the tethered dsDNA (long) to ssDNA (short), displacing the bead opposite the flow. **d**, Representative synthesis trajectories upon encountering Tus–TerB oriented to the P (P<sub>TerB</sub>, left) or NP (NP<sub>TerB</sub>, right) faces. kbp, kilobase pairs. **e**, Percentages of forks that bypassed, transiently or fully stopped at the P or NP face. Error bars correspond to standard deviations of binomial distributions; N = 60, 64 and 37 for NP<sub>TerB</sub>

(–Tus), NP<sub>TerB</sub> (+Tus) and P<sub>TerB</sub> (+Tus), respectively. **f**, Effect of Tus concentration on arrest activity at the NP face. Tus was present continuously with the replication proteins. Washing excess DNA-unbound Tus (80 nM) before introduction of replication proteins resulted in 38% stoppage. **g**, Rate dependence of replication stalling at the NP face. Rate distributions of events that bypassed (grey; N = 33) or stopped/restarted (blue bars; N = 31) were fit with Gaussian distributions. Fit lines are shown; the uncertainty corresponds to the standard error. **h**, Percentages of forks that bypassed, transiently or fully stopped at Tus bound to the NP<sub>TerB</sub> site containing a bubbled-DNA structure in place of base pairs 3–7 of TerB, while keeping C6 (5-mismatch C6–NP<sub>TerB</sub>). Error bars correspond to standard deviations of binomial distributions; N = 14 and 27 in absence and presence of Tus, respectively.

<sup>1</sup>Division of Biological and Environmental Sciences and Engineering, King Abdullah University of Science and Technology, Thuwal 23955, Saudi Arabia. <sup>2</sup>Centre for Medical & Molecular Bioscience, Illawarra Health & Medical Research Institute and University of Wollongong, New South Wales 2522, Australia.

oriented groups<sup>2,3</sup>, and each of them is tightly bound to the monomeric protein Tus<sup>6,10</sup>. The lack of symmetry in *Ter* sequences fixes the orientation of the Tus–*Ter* complex such that forks are blocked at its non-permissive (NP) face, but allowed to pass from the permissive (P) end<sup>6,11</sup>. The two *Ter* clusters thus form a trap from which the first arriving fork can enter but not leave, awaiting arrival of the other<sup>2–5</sup>.





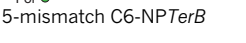
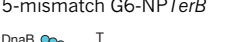


The mechanism determining polarity of Tus–*Ter* action serves as a model for communication between replication forks and double-stranded (ds) DNA-binding proteins, but it is also controversial. Strand separation by the DnaB helicase at the NP face can engineer a new structure, the ‘locked’ complex of the mousetrap model (Fig. 1b)<sup>6</sup>. Cytosine(6) of *Ter* flips out of the DNA helix to form new interactions in a pocket on Tus that markedly prolongs the lifetime (>40-fold) of the Tus–*Ter* complex, protecting the central interactions from the trailing polymerase. Conversely, strand separation at the P face rapidly dissociates Tus.

Despite the stability of the locked complex *in vitro*, at any sampling time in reporter plasmids *in vivo*, even when Tus is overproduced, ~50% of forks moving towards the NP face displace Tus<sup>8,9</sup>. This could be due either to the fork block being transient or to its low efficiency of formation. The  $K_D$  of the Tus–*TerB* locked complex is only threefold lower than Tus–ds*TerB* while its lifetime is much longer<sup>6</sup>, so we tested the hypothesis that the efficiency of lock formation is kinetically controlled, that is, that NP fork-arrest efficiency is determined by competition between lock formation and Tus displacement, dependent on the rate of fork approach. Inherent inefficiency of fork arrest would also explain the presence of backup *Ter* sites in the chromosome.

We used single-molecule imaging to monitor the fate of the *E. coli* leading-strand replisome as it approaches Tus–*TerB* from either direction. Real-time synthesis trajectories were derived from multiplexed arrays by monitoring the length of individual DNA molecules<sup>12–14</sup>. The forked primer–template DNAs, each with a single *TerB* site 3.6 kilobase (kb) from the site of fork assembly, were tethered between the surface of a coverslip and a magnetic bead (Fig. 1c and Extended Data Fig. 1) and extended by a laminar flow exerting a 2.6 pN drag force on the beads. The trajectories (Fig. 1d) show DNA shortening through its conversion from ds (long) to single-stranded (short) during leading-strand synthesis. The position of *TerB* could be defined to  $\pm 0.1$  kb under this force regime (see Methods). Consistent with previous single-molecule studies of DNA replication<sup>12–17</sup>, rates of DNA synthesis vary among replisomes (Extended Data Fig. 2), reflecting the *in vivo* situation<sup>18</sup>. The source of this heterogeneity is unknown (Supplementary Discussion 1).

In the absence of Tus,  $5 \pm 3\%$  of forks that reached the *TerB* site stopped there by chance (Table 1, Fig. 1e, Extended Data Fig. 3a, b and Supplementary Discussion 2). With Tus–*TerB* oriented with its P face towards the fork (P*TerB*), this frequency increased to  $11 \pm 5\%$  (including the 5% random stoppage), presumably owing to forks encountering a strong protein–DNA roadblock (Fig. 1e and Supplementary Discussion 3). Transient stoppage followed by resumption of synthesis occurred in 5% of trajectories, and in the remaining 84%, replication forks displaced Tus and continued synthesis without stopping, even transiently (Fig. 1d, e). The average rate of DNA synthesis was otherwise unaffected by Tus (Extended Data Fig. 3c).

**Table 1 | Fate of replisomes and fork rate dependencies of events at Tus-bound *Ter* sites**

<i>Ter</i> 5'-AATAAGTATGTTGTAACAAAGT P NP TTATTGATACACATGATTCA-5'	Tus	Stop (%)	Bypass (%)	Restart (%)	Pause time (s)	Stop/restart rate (bp s <sup>-1</sup> )	Bypass rate (bp s <sup>-1</sup> )
	WT Tus	11 ± 5	84 ± 6	5 ± 4	31 ± 24	330 ± 30	1,300 ± 110 (1,250 ± 120)
P <i>TerB</i>	No Tus	5 ± 3	95 ± 3	0	–	800 ± 100	1,160 ± 70 (930 ± 70)
	WT Tus	45 ± 6	52 ± 6	3 ± 2	146 ± 31	890 ± 70 (840 ± 40)	1,690 ± 100 (1,700 ± 140)
Pol	H144A	27 ± 8	55 ± 9	18 ± 7	33 ± 5	840 ± 120 (740 ± 10)	1,520 ± 130 (1,520 ± 120)
NP <i>TerB</i> [GC(6)-NP <i>TerB</i> ]	R198A*	5 ± 5	77 ± 9	18 ± 8	14 ± 4	400 ± 40	1,300 ± 140 (1,310 ± 70)
	WT Tus	8 ± 4	47 ± 8	45 ± 8	29 ± 6 (37 ± 6)	820 ± 70 (720 ± 80)	1,810 ± 110 (1,740 ± 60)
CG(6)-NP <i>TerB</i>	WT Tus†	18 ± 9	41 ± 11	41 ± 11	56 ± 13	800 ± 80 (790 ± 70)	1,820 ± 180
	WT Tus	7 ± 4	73 ± 7	20 ± 6	22 ± 5 (24 ± 2)	380 ± 60 (360 ± 40)	1,350 ± 110 (1,310 ± 50)
TA(6)-NP <i>TerB</i>	WT Tus	89 ± 6	11 ± 6	0	–	1160 ± 130 (1030 ± 130)	1,230 ± 310
	R198A*	86 ± 7	14 ± 7	0	–	1130 ± 120 (950 ± 170)	1,200 ± 270
5-mismatch C6-NP <i>TerB</i>	WT Tus	8 ± 5	15 ± 7	77 ± 8	111 ± 18 (177 ± 20)	1130 ± 140 (960 ± 170)	1,230 ± 170
	WT Tus	32 ± 9	68 ± 9	0	–	390 ± 70 (360 ± 40)	1,480 ± 140 (1,320 ± 60)
5-mismatch G6-NP <i>TerB</i>	WT Tus	41 ± 9	59 ± 9	0	–	880 ± 110 (860 ± 90)	1,480 ± 120 (1,550 ± 110)
	WT Tus	22 ± 7	75 ± 7	3 ± 3	186	400 ± 60 (310 ± 70)	1,260 ± 100 (1,260 ± 60)
Swapped F4n GC(6)-NP <i>TerB</i>	WT Tus	19 ± 6	70 ± 8	11 ± 5	180 ± 26	400 ± 60 (350 ± 20)	1,320 ± 100 (1,300 ± 60)
	WT Tus	–	–	–	–	–	–
Swapped F5n GC(6)-NP <i>TerB</i>	NP <i>TerH</i>	–	–	–	–	–	–

The *TerB* site and its alterations are depicted in cartoon with the native nucleotides in black except that C6 is in yellow and substituted nucleotides are in magenta. The directionality of the replication fork is shown with the strands on which Pol III holoenzyme and DnaB translocate. The nucleotides in native *TerH* that differ from *TerB* are also shown in magenta. The sequences of oligonucleotides used to assemble the variants of *TerB* substrates are given in Extended Data Fig. 1b.

Stop, bypass and restart events were quantified as a percentage of all events that reached or bypassed *TerB* (see Methods). Uncertainties correspond to standard deviations of binomial distributions. The mean of the rates and pause durations are shown either as their arithmetic averages or, in parentheses, by fitting their histograms with Gaussian or exponential decay distributions, respectively (see Fig. 2a, c and Methods). The uncertainties correspond to standard errors. Each experimental condition represents results from three or four technical replicates and the number of derived molecules (*N*) is specified in the corresponding figures.

\*Concentration of Tus(R198A) was 250 nM versus the standard 80 nM.

†Reaction from which the restart proteins PriA, PriB and DnaT were omitted.

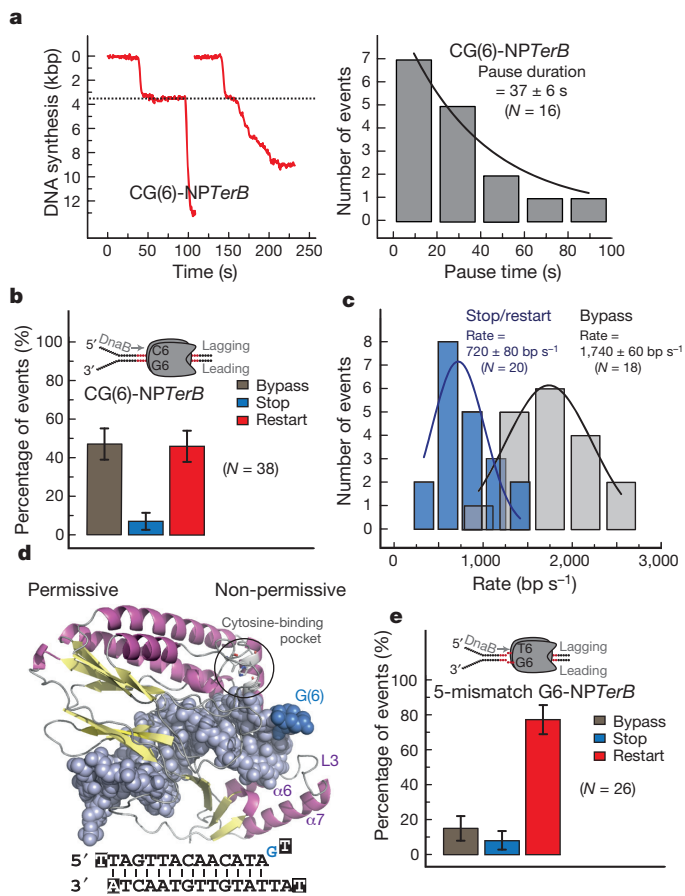
In contrast, when the fork approached the NP face (NP<sub>TerB</sub>), permanent stoppage ( $9.7 \pm 1$  min) of DNA synthesis occurred in 45% of trajectories, and restart in only 3% (Fig. 1d, e and Extended Data Fig. 2a, b). The remaining 52% showed no sign even of transient stoppage. All *TerB* sites were Tus-bound under our experimental conditions (Fig. 1f), indicating they have an inherently low efficiency of fork arrest. We are thus able for the first time to distinguish between the two different mechanisms that could explain the *in vivo* data<sup>8,9</sup>.

Fork arrest is attenuated *in vivo* by DNA supercoiling, suggesting that it is affected by the rate of strand separation<sup>8</sup>. To test this proposition, we separated the trajectories that showed full or transient stoppage from those that did not and found the rate of DNA synthesis and fork bypass were correlated ( $r = 0.62$ ; Extended Data Fig. 4a); fast forks were arrested less often than slower ones. In fact, there was a twofold difference in average rates of synthesis at forks that stopped and those that bypassed *TerB* (Fig. 1g). DNA synthesis at individual forks before stoppage at *TerB*, or in full trajectories where they bypassed it, progressed at nearly constant rates under our spatial and temporal resolution (Extended Data Fig. 4b–e). As we showed previously<sup>14</sup>, the overall average rate reproduces the average *in vivo* rate ( $\sim 950$  bp s<sup>-1</sup>)<sup>18</sup>. This underscores the significance of our ability to achieve the *in vivo* rate of DNA synthesis to reproduce the  $\sim 50\%$  efficiency of fork arrest *in vivo*.

The rate dependence of stoppage supports the hypothesis that strand separation competes with inefficient C6 flipping. To demonstrate this, we pre-formed the locked complex before replisome assembly using *TerB* with a mismatched bubble in place of base pairs 3–7 while keeping an unpaired C6 (Fig. 1h and Table 1)<sup>6</sup>. The yield of fork arrest increased to 89%; thus, once the lock is established, it is a very effective fork block.

We next interrogated the role of lock formation with C6-defective *TerB* mutants<sup>6</sup>. Surprisingly, the GC(6) to CG substitution did not lead to  $\sim 95\%$  bypass. Instead, it resulted in transient (for  $37 \pm 6$  s) rather than permanent blockage, again in  $\sim 50\%$  of trajectories (Fig. 2a, b and Extended Data Fig. 2c). Moreover, the fork rate dependence of pausing was similar to the normal lock (Fig. 2c). DnaB remained at the fork during transient arrest since DNA synthesis could restart in the absence of helicase reloading proteins (Table 1). The crystal structure of a Tus complex with a forked *Ter* containing an unpaired G replacing C6 showed that the substituted G6 base neither bound in the cytosine pocket nor formed any new specific interaction with Tus (Fig. 2d and Extended Data Table 1). This remained the case even when the fork was extended to also disrupt the TA(7) base pair (Extended Data Fig. 5a–c). Thus the fork-rate-dependent step producing transient stoppage must precede engagement of C6 in its binding pocket.

In the Tus crystal structures, the  $\alpha 6/L3/\alpha 7$  region has extensive interactions with the lagging strand (Fig. 1b) before and after C6 lock formation<sup>6,11</sup>, providing a paradox about how the lagging-strand-translocating DnaB in fast-approaching replisomes disrupts these interactions without even pausing. The main sequence-specific contact this region makes with the first 6 bp of ds*Ter* are via Arg198 in L3 with the A5 and G6 bases on the lagging strand and T5 on the leading strand (Extended Data Fig. 5e)<sup>11</sup>, but these interactions are not present in the locked complex, where Arg198 makes a new salt bridge to the phosphate between lagging strand nucleotides 6 and 7 (Extended Data Fig. 5a)<sup>6</sup>. We suggest that the Arg198 side chain forms transient interactions with G6, the TA(5) base pair and the lagging-strand phosphate, holding the two DNA strands together before strand separation reaches GC(6) (Extended Data Fig. 5e). Moreover, comparison of the structures of Tus with the wild type<sup>11</sup> and CG(6) mutant ds*Ter* sites (Extended Data Table 1) suggests rearrangement of lagging strand interactions with Arg198 (Extended Data Fig. 5e, f). We propose that Arg198–DNA contacts rearrange substantially during strand separation. This provides a window of opportunity for the fast-moving DnaB to break into the Tus–*Ter* central interactions before Arg198 rearrangement or C6 base flipping occurs.



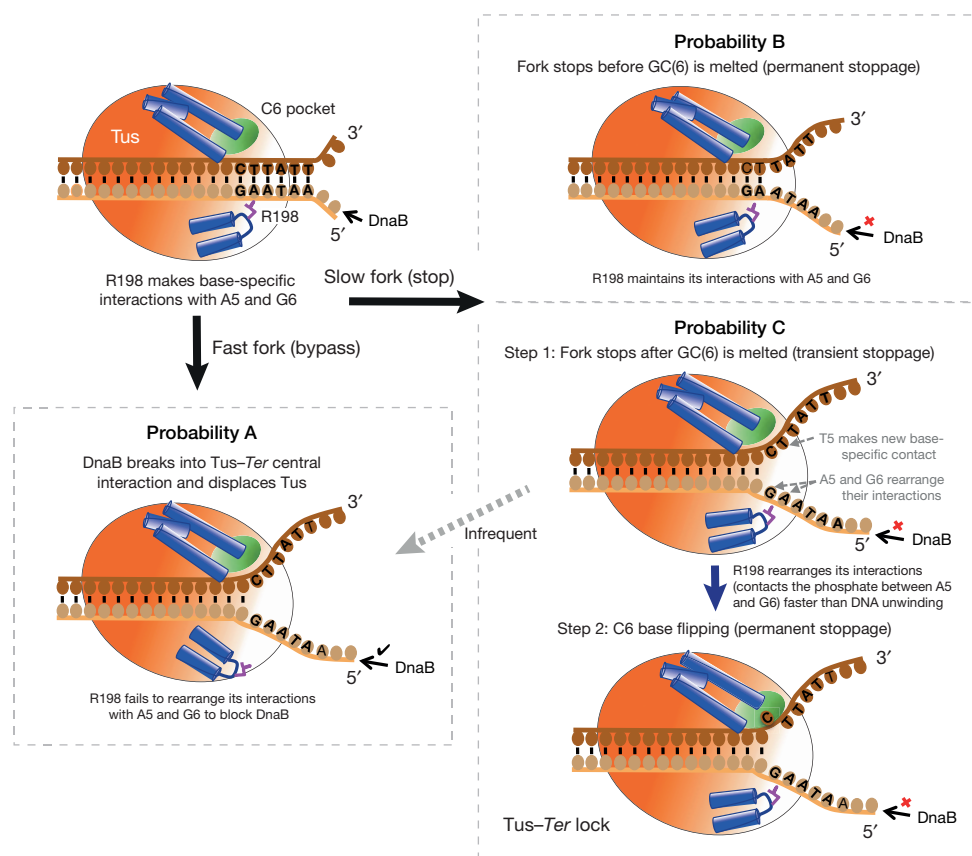
**Figure 2 | Characterization of transient stoppage of the replication fork at the non-permissive face of Tus–*TerB* before C6 base flipping.**

**a**, Representative trajectories of restart of DNA synthesis after transient stoppage at a Tus-bound *TerB* site where the GC(6) base pair was swapped to CG(6) (CG(6)-NP<sub>TerB</sub>). The distribution of pause durations was fit with a single exponential decay; the fit line is shown and uncertainty corresponds to the standard error ( $N = 16$ ). **b**, Percentages of the populations of replication forks that bypassed, transiently stopped or fully stopped at CG(6)-NP<sub>TerB</sub>. Error bars correspond to standard deviations of binomial distributions ( $N = 38$ ). **c**, Rate dependence of replication restart at CG(6)-NP<sub>TerB</sub>. The rate distribution of leading-strand synthesis for events that bypassed (grey;  $N = 18$ ) or stopped/restarted (blue bars;  $N = 20$ ) at CG(6)-NP<sub>TerB</sub> are presented as in Fig. 1g. **d**, Crystal structure of Tus with a forked *Ter* sequence that has a substituted G base in the C6 position in the locked complex (see also Extended Data Fig. 5b). The G base, highlighted in blue, was neither docked into the cytosine-binding pocket nor forming any new interactions with Tus. Highlighted nucleotides (at bottom) were not visible in the structure. **e**, Fates of replication forks at Tus bound to the NP face of a *TerB* site containing a bubbled-DNA structure in place of base pairs 3–7 in *TerB* and with G replacing C6 (5-mismatch G6-NP<sub>TerB</sub>). Error bars correspond to standard deviations of binomial distributions ( $N = 26$ ).

To test this proposition, we used a bubble substrate with altered C6 (5-mismatch G6-NP<sub>TerB</sub>; see Table 1) to eliminate lock formation<sup>6</sup> but allow rearrangement of interactions on the separated strands before arrival of DnaB. We observed efficient transient stoppage that reached 77% with a long duration of  $177 \pm 20$  s (Fig. 2e; Extended Data Fig. 6). The fivefold increase in pause duration compared to CG(6)-NP<sub>TerB</sub> (Table 1) is probably owing to the interactions of the unpaired seventh nucleotide as in the locked complex structure<sup>6</sup>. Thus, strand separation beyond GC(6) in the absence of the C6 lock would impose only transient fork stoppage.

We next altered A5 and T5 alone and in the context of the first 5 bp of *TerB* (Table 1); this resulted in the largest decrease in yield of stoppage and shift in the rate-dependence of arrest to lower values,





**Figure 3 | Model of Tus–Ter polar arrest activity at the non-permissive face.** Prior to strand separation, Arg198 makes base-specific contacts with A5 and G6 on the lagging strand to protect Tus–Ter central interactions from the DnaB helicase. After separation of the first six base pairs, Arg198 maintains contacts with the lagging strand by rearranging its interactions to make a new salt bridge to the phosphate between A5 and G6 and a new unidentified base-specific interaction is induced with T5. Competition between rates of strand separation and rearrangement of Arg198 interactions determines Tus–Ter efficiency.

Faster moving forks have higher probability to separate GC(6) before rearrangement of Arg198 interactions, leading to effective displacement of Tus (probability A). The slower forks are either stopped permanently before or during GC(6) melting (probability B) or transiently if GC(6) is melted and Arg198 succeeds in rearranging its interactions (probability C, step 1). The C(6) mousetrap acts as a terminal step that is enabled by the transient stoppage to impose permanent fork arrest (probability C, step 2).

underscoring that A5 and/or T5 are the primary contributors in this region to the rate dependence (Extended Data Fig. 7a–c and Supplementary Discussion 4). AT(5) is conserved in strong *Ter* sites but not at weaker ones like *TerH*<sup>9</sup>. We also found that the NP Tus–*TerH* complex stops the forks with a similarly low rate dependence to TA(5)–NP*TerB* (Table 1). However, one-third of the stopped forks restarted synthesis after a pause of  $180 \pm 26$  s (Extended Data Fig. 7g, h), which we attribute to other alterations in *TerH* that weaken its binding in its locked form<sup>19</sup>.

Substitution of wild-type GC(6) with CG has a modest effect on Tus binding to ds*TerB* in comparison to an AT or TA<sup>3</sup>. We observed that, relative to CG(6)–NP*TerB*, the TA substitution resulted in transient fork stoppage with decreased yield (Table 1 and see Fig. 2a–c; Extended Data Fig. 7d–f), demonstrating the importance of the specific interaction of Tus with the native G6 for transient stoppage.

We then altered Arg198 itself. The R198A mutant interacts with ds*TerB* with a 140-fold increased  $K_D$ , but only a twofold shorter lifetime<sup>20</sup>. We showed by surface plasmon resonance (SPR) that Tus(R198A) can form a lock (Extended Data Fig. 8a–d), but it was very defective in fork arrest (Table 1); stoppage was inefficient (18%) and transient (pauses of  $14 \pm 4$  s;  $N = 4$ ). Nevertheless, preforming the locked complex with R198A on the 5-mismatch C6–NP*TerB* substrate restored efficient stoppage (Table 1), consistent with lock formation. These results suggest that C6 flipping cannot occur unless Arg198 interactions slow down or transiently stop the fork beforehand.

So we have by now revealed two separate processes, one leading to a transient stoppage preceding but probably on the pathway to C6 lock formation, and one that leads directly to bypass and Tus dissociation. Previous results have suggested the operation of an uncharacterized C6-lock-independent arrest mechanism<sup>3,21</sup>. Our study shows that this mechanism must be invoked before or as GC(6) is melted, because permanent stoppage was not achieved when the GC(6) was melted in the absence of the C6 lock (Fig. 2e). To explore whether the interactions of Arg198 with AT(5) and G6 contribute directly to this alternate mechanism, we maintained these interactions using the *TerB* sequence and deactivated the C6 lock using Tus(H144A), the key residue in the binding pocket. This mutation completely eliminated lock formation (SPR in Extended Data Fig. 8e, f; X-ray structure in Extended Data Fig. 5d). However, we still observed a high level (27%) of permanent fork arrest, confirming existence of a lock-independent process leading to permanent stoppage. There were also significant restarts (18%; Table 1) after short pauses ( $33 \pm 5$  s;  $N = 6$ ). Pausing must result from a mechanism additional to permanent arrest, since restarts would otherwise be randomly distributed over the full 10 min period of observation. The rate-dependence of arrest was similar to wild-type Tus (Table 1 and Extended Data Fig. 9).

Collectively, our results show that interactions of Arg198 of Tus with G6, A5 and/or T5 act to protect Tus–Ter central interactions from the first arriving DnaB. Nevertheless, these gatekeeping interactions are dynamic during separation of the first 6 bp and their rearrangement

occurs in competition with strand separation. We suggest that faster forks have higher probability to separate GC(6) before rearrangement of Arg198 interactions, displacing Tus without pausing (Fig. 3, probability A). Slower forks are either stopped permanently before GC(6) melting (probability B) or transiently if GC(6) is melted and Arg198 succeeds in rearranging (probability C). The inefficient C6 mousetrap is a terminal step, enabled by transient stoppage to impose permanent fork arrest (probability C). These results provide an explanation of why, in helicase assays, the slowly moving DnaB ( $35\text{--}390\text{ bp s}^{-1}$ )<sup>22,23</sup> is efficiently stopped at the NP face without requiring C6 flipping<sup>21</sup>.

Thus, we refine the mousetrap model and redefine the efficiency of Tus–Ter polar arrest to depend on collective contributions of intrinsic affinity of Tus for Ter, stability of the flipped C6 in its binding pocket, and rate-dependent induction of fork stoppage that fully or temporarily protects Tus–Ter central interactions from DnaB. Our observations also raise a question about how weaker Ter sites evolved to block slower forks (Supplementary Discussion 5). The encounter of dsDNA-binding proteins with motor proteins like helicases and polymerases is a common feature in replication, repair, recombination and transcription and where conflict among these processes arises (reviewed in ref. 24). We show for the first time that intrinsic heterogeneity in rates of individual molecular motors can have different biological outcomes as they communicate with dsDNA-binding proteins and other barriers (Supplementary Discussion 6).

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 21 January; accepted 26 June 2015.

Published online 31 August 2015.

1. Dalgaard, J. Z. *et al.* Random and site-specific replication termination. *Methods Mol. Biol.* **521**, 35–53 (2009).
2. Hill, T. M., Henson, J. M. & Kuempel, P. L. The terminus region of the *Escherichia coli* chromosome contains two separate loci that exhibit polar inhibition of replication. *Proc. Natl Acad. Sci. USA* **84**, 1754–1758 (1987).
3. Coskun-Ari, F. F. & Hill, T. M. Sequence-specific interactions in the Tus–Ter complex and the effect of base pair substitutions on arrest of DNA replication in *Escherichia coli*. *J. Biol. Chem.* **272**, 26448–26456 (1997).
4. Neylon, C., Kralicek, A. V., Hill, T. M. & Dixon, N. E. Replication termination in *Escherichia coli*: structure and antihelicase activity of the Tus–Ter complex. *Microbiol. Mol. Biol. Rev.* **69**, 501–526 (2005).
5. Duggin, I. G., Wake, R. G., Bell, S. D. & Hill, T. M. The replication fork trap and termination of chromosome replication. *Mol. Microbiol.* **70**, 1323–1333 (2008).
6. Mulcair, M. D. *et al.* A molecular mousetrap determines polarity of termination of DNA replication in *E. coli*. *Cell* **125**, 1309–1319 (2006).
7. Kaplan, D. L. & Bastia, D. Mechanisms of polar arrest of a replication fork. *Mol. Microbiol.* **72**, 279–285 (2009).
8. Valjavec-Gratian, M., Henderson, T. A. & Hill, T. M. Tus-mediated arrest of DNA replication in *Escherichia coli* is modulated by DNA supercoiling. *Mol. Microbiol.* **58**, 758–773 (2005).
9. Duggin, I. G. & Bell, S. D. Termination structures in the *Escherichia coli* chromosome replication fork trap. *J. Mol. Biol.* **387**, 532–539 (2009).
10. Coskun-Ari, F. F., Skokotas, A., Moe, G. R. & Hill, T. M. Biophysical characteristics of Tus, the replication arrest protein of *Escherichia coli*. *J. Biol. Chem.* **269**, 4027–4034 (1994).
11. Kamada, K., Horiuchi, T., Ohsumi, K., Shimamoto, N. & Morikawa, K. Structure of a replication-terminator protein complexed with DNA. *Nature* **383**, 598–603 (1996).
12. Lee, J. B. *et al.* DNA primase acts as a molecular brake in DNA replication. *Nature* **439**, 621–624 (2006).
13. Tanner, N. A. *et al.* Single-molecule studies of fork dynamics in *Escherichia coli* DNA replication. *Nat. Struct. Mol. Biol.* **15**, 170–176 (2008).
14. Jergic, S. *et al.* A direct proofreader-clamp interaction stabilizes the Pol III replisome in the polymerization mode. *EMBO J.* **32**, 1322–1333 (2013).
15. Tanner, N. A. *et al.* Real-time single-molecule observation of rolling-circle DNA replication. *Nucleic Acids Res.* **37**, e27 (2009).
16. Hamdan, S. M., Loparo, J. J., Takahashi, M., Richardson, C. C. & van Oijen, A. M. Dynamics of DNA replication loops reveal temporal control of lagging-strand synthesis. *Nature* **457**, 336–339 (2009).
17. Yao, N. Y., Georgescu, R. E., Finkelstein, J. & O'Donnell, M. E. Single-molecule analysis reveals that the lagging strand increases replisome processivity but slows replication fork progression. *Proc. Natl Acad. Sci. USA* **106**, 13236–13241 (2009).
18. Pham, T. M. *et al.* A single-molecule approach to DNA replication in *Escherichia coli* cells demonstrated that DNA polymerase III is a major determinant of fork speed. *Mol. Microbiol.* **90**, 584–596 (2013).
19. Moreau, M. J. & Schaeffer, P. M. Differential Tus–Ter binding and lock formation: implications for DNA replication termination in *Escherichia coli*. *Mol. Biosyst.* **8**, 2783–2791 (2012).
20. Neylon, C. *et al.* Interaction of the *Escherichia coli* replication terminator protein (Tus) with DNA: a model derived from DNA-binding studies of mutant proteins by surface plasmon resonance. *Biochemistry* **39**, 11989–11999 (2000).
21. Bastia, D. *et al.* Replication termination mechanism as revealed by Tus-mediated polar arrest of a sliding helicase. *Proc. Natl Acad. Sci. USA* **105**, 12831–12836 (2008).
22. Ribbeck, N., Kaplan, D. L., Bruck, I. & Saleh, O. A. DnaB helicase activity is modulated by DNA geometry and force. *Biophys. J.* **99**, 2170–2179 (2010).
23. Kim, S., Dallmann, H. G., McHenry, C. S. & Mariani, K. J. Coupling of a replicative polymerase and helicase: a  $\tau$ -DnaB interaction mediates rapid replication fork movement. *Cell* **84**, 643–650 (1996).
24. Finkelstein, I. J. & Greene, E. C. Molecular traffic jams on DNA. *Annu. Rev. Biophys.* **42**, 241–263 (2013).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank A. van Oijen for critical comments and the groups of N. Dekker and S. Patel for helpful discussions. This research was supported by the King Abdullah University of Science and Technology through core funding (to S.M.H.) and a Faculty Initiated Collaborative Award (to S.M.H. and N.E.D.), and by the Australian Research Council (DP0877658 to N.E.D. and A.J.O.; DP0984797 to N.E.D.), including an Australian Professorial Fellowship to N.E.D. and a Future Fellowship (FT0990287) to A.J.O. X-ray crystallographic data were collected at the Australian Synchrotron, Victoria, Australia.

**Author Contributions** M.M.E. designed and carried out the single-molecule replication assays; M.M.E., M.A.S. and M.T. established the single-molecule replication assays; S.J. designed and carried out SPR measurements; S.J. and Z.-Q.X. isolated proteins; Z.-Q.X. and A.J.O. crystallized complexes, collected X-ray data and refined crystal structures. M.M.E., S.J., N.E.D. and S.M.H. designed the research and wrote the article. All authors analysed the data, discussed the results and commented on the manuscript.

**Author Information** Atomic coordinates and structure factors for the reported crystal structures have been deposited at the Protein Data Bank under accession codes 4XR0 (Tus–UGLT fork), 4XR1 (Tus–TGTA fork), 4XR2 (H144A–WT fork) and 4XR3 (Tus–UGLC). Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to N.E.D. ([nickd@uow.edu.au](mailto:nickd@uow.edu.au)) or S.M.H. ([samir.hamdan@kaust.edu.sa](mailto:samir.hamdan@kaust.edu.sa)).

## METHODS

No statistical methods were used to predetermine sample size, the experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

**Protein expression and purification.** Described methods were used to prepare N-terminally His<sub>6</sub>-tagged Tus<sup>20</sup> and its mutant derivatives Tus(R198A)<sup>20</sup> and Tus(H144A)<sup>6</sup>, as well as the following *E. coli* DNA replication proteins: the  $\beta_2$  sliding clamp<sup>25</sup>, the Pol III  $\tau_3\delta\delta'\chi\psi$  clamp loader and  $\alpha\epsilon\theta$  core<sup>13</sup>, the DnaB<sub>6</sub>(DnaC)<sub>6</sub> helicase/loader complex<sup>14</sup>, and the fork restart proteins PriA, PriB and DnaT<sup>26</sup>.

**Crystallization of Tus–Ter complexes and data collection.** Four crystal structures of Tus–Ter complexes are reported (Extended Data Table 1; the oligonucleotide sequences and proteins used are given in Extended Data Fig. 5). All complexes (finally at 4–5 mg ml<sup>−1</sup> protein) were prepared with a slight excess of DNA in 10 mM Bis-Tris, pH 6.5, 1 mM EDTA, 2 mM dithiothreitol, and excess DNA was removed by using a centrifugal ultrafiltration device, as described previously<sup>6</sup>. Crystals were grown using the vapour diffusion (hanging drop) method at 23 °C. The protein–DNA complex (3  $\mu$ l) was mixed with an optimized reservoir solution (3  $\mu$ l) consisting of 8–12% PEG 3350, 0.1–0.2 M NaI, 50 mM Bis-Tris, pH 6.2–6.8. Crystals appeared after 2 days and reached maximum size after 10 days. The pH was measured using 1 M stock solutions of buffer before addition to the reservoir. Optimized reservoir solutions for the four crystals contained: for Tus–UGLT fork (forked Ter with C6 to G change), 12% PEG 3350, 0.2 M NaI, 50 mM Bis-Tris pH 6.2; for Tus–TGTA fork (forked Ter with C6 to G change and the fork extended to position 7), 9% PEG 3350, 0.2 M NaI, 50 mM Bis-Tris pH 6.8; for Tus(H144A)–WT fork ('wild-type' forked Ter with Tus(H144A)), 8% PEG 3350, 0.1 M NaI, 50 mM Bis-Tris pH 6.2; for Tus–UGLC (dsTerA with GC(6) to CG flip), 12% PEG 3350, 0.2 M NaI, 50 mM Bis-Tris pH 6.5.

All X-ray data were collected at the Australian Synchrotron beamline MX-1 (X-ray wavelength, 0.95370 Å) using an Oxford cryostream to maintain the crystal temperature at 100 K. Prior to cooling, crystals were transferred stepwise into artificial mother liquors finally containing 15% (v/v) MPD (2-methyl-2,4-pentandiol) in 3% increments of MPD (3 min per step). Data were collected using an ADSC Quantum 210r area detector, using BLU-ICE for remote data acquisition and processing<sup>27</sup>. Data reduction and scaling was achieved with the HKL2000 package<sup>28</sup>.

**Structure determination and refinement.** All structures were solved by molecular replacement in MOLREP<sup>29</sup> using a previously solved Tus–Ter lock (PDB code: 2I06) or Tus–TerA structure (2I05) as starting model. REFMAC<sup>30,31</sup> was used for structure refinement and calculation of map weighting factors. COOT<sup>32</sup> was used to interpret electron density maps and for model building. Figures were prepared using PyMOL<sup>33</sup>.

**Assessment of Tus–TerB interactions by surface plasmon resonance (SPR).** Methods were essentially as used previously<sup>6,20</sup>, except that all experiments were carried out at 20 °C (instead of 25 °C) and a 6  $\times$  6 multiplex BioRad ProteOn XPR-36 system was used instead of a Biacore 2000 instrument; dissociation rate constants ( $k_d$ ) of Tus proteins from immobilized biotinylated TerB showed an unusual temperature dependence (high activation energy), which accounts for lower values of  $k_d$  and the dissociation constant ( $K_D$ ) compared to previously reported values (where they are available<sup>6,20</sup>).

All measurements used SPR buffer (50 mM Tris pH 7.6, 250 mM KCl, 0.25 mM EDTA, 0.5 mM dithiothreitol, 0.005% surfactant P<sub>20</sub>), with a ProteOn NLC (neutravidin-coated) sensor chip for immobilization of 5'-biotinylated TerB oligodeoxynucleotides (oligos). These were either (1) 5'-bio-(pD)<sub>10</sub>-ATAAGTATGT TGTAACAAAG, oligo-1, or (2) 5'-bio-(pD)<sub>10</sub>-GGGGCTATGTTGTAACAAAG, oligo-2, each containing a 10-unit abasic deoxyribosephosphate spacer (pD)<sub>10</sub> to move the TerB molecule away from the chip surface<sup>20</sup>, as well as a common lagging strand TerB sequence (underlined)<sup>20</sup>. Hybridization of oligo-3: 5'-CTTTAGTTACAACATACTTAT (C6 of Ter in bold) to oligo-1 produces a full dsTerB site, while its hybridization to oligo-2 produces a forked Ter where C6 is unpaired and exposed (mismatched sequences in oligos-2 and -3 are in italics)<sup>6</sup>.

All 36 interaction spots of the sensor chip were activated with three sequential injections of 1 M NaCl, 50 mM NaOH across six vertical (ligand) flow paths (40 s each at 40  $\mu$ l min<sup>−1</sup>) and six horizontal (analyte) flow paths (40 s each at 100  $\mu$ l min<sup>−1</sup>). The surface was further stabilized by two injections of 1 M MgCl<sub>2</sub> in each direction, with the same contact times and flow rates. Oligos-1 and -2 were diluted to 200 nM in SPR buffer and immobilized separately onto the six interaction spots of the vertical flow path (100  $\mu$ l min<sup>−1</sup> for 15 s). The chip was then rotated 90° and simultaneous assembly of dsTerB and forked Ter templates<sup>6</sup> on the chip surface was achieved by hybridization of oligo-3 (300 nM), made to flow across all six horizontal (analyte) channels at 25  $\mu$ l min<sup>−1</sup> for 400 s. The sensorgram verified that hybridization went to completion. After subsequent injection of a concentration series of Tus, the surface was regenerated; remaining proteins and

hybridized DNAs were removed by two injections of 1 M NaCl, 50 mM NaOH over the six analyte channels at 50  $\mu$ l min<sup>−1</sup> for 40 s, followed by re-hybridization of oligo-3 as above. Measured stoichiometries of Tus binding to both templates were close to 1:1 at saturation, as reported previously<sup>6,20</sup>.

Tus, Tus(R198A) and Tus(H144A) interactions with TerB and forked Ter templates were carried out by sequential injections in the analyte direction of one or two appropriate concentration series in SPR buffer (zero and five concentrations of serially diluted samples) at 40  $\mu$ l min<sup>−1</sup> for 300 s, followed by dissociation in the same buffer over 2,000 s. The final sensorgrams were interspot and unmodified ligand flow path subtracted using ProteOn Manager Software (v. 3.1.0.6) and then zero subtracted and normalized based on the highest response of hybridized oligo within the discrete ligand flow path using BIAevaluation software (v. 4.0.1; Biacore AB, Sweden). Equilibrium (dissociation constant,  $K_D$ ) and kinetic (rate constants,  $k_a$  and  $k_d$ ) parameters for the binding of Tus proteins to the Ter fragments were determined by global (simultaneous) fitting of at least five sensorgrams per measured interaction from the optimized concentration range using BIAevaluation software and the appropriate interaction model(s): (Langmuir) 1:1 binding with mass transfer model (LMT, for Tus–dsTerB interaction; as previously done<sup>20</sup>), (Langmuir) 1:1 binding model (L, for Tus– and Tus(R198A)–forked Ter), and 1:1 steady state affinity (LSS) and heterogeneous ligand-parallel reactions (HLPR) binding models for fitting sensorgrams that reached an equilibrium response (Tus(R198A)–dsTerB, Tus(H144A)–dsTerB and Tus(H144A)–forked Ter).

Global best fits were used when LSS and HLPR models were used. When L and LMT models were used, the fitting was constrained by setting the  $R_{max}$  to a global constant value (response at saturation of ligand binding sites was set to 700 response units (RU) for bindings to dsTerB and 775 RU for forked Ter). These values, calculated theoretically as a product of the highest measured response of hybridized oligo-3 (molecular weight 6,354; used as a normalization unit) onto oligo-1 (120 RU) and oligo-2 (134 RU) and the factor 5.8 (molecular weight of His<sub>6</sub>–Tus/molecular weight of hybridized oligo-3 = 36,737/6,354), were compared with experimentally determined values obtained by flowing Tus at a saturating concentration (1.024  $\mu$ M) over the two DNA templates (not shown). In addition, due to slow dissociation, experimentally determined  $k_d$  values for Tus– and Tus(R198A)–forked Ter interactions using the L model were assessed by comparison with the  $k_d$  values determined from the experiment where dissociation was monitored over 50,000 s (not shown). To generate as reliable as possible values for kinetic parameters using the HLPR model,  $k_a$  and  $k_d$  were estimated in the first approximation based on complete association phase and only an initial phase of dissociation where the rate of change is the greatest. These obtained values of kinetic parameters were sometimes used as initial iterative values; otherwise, the iterations could slip into local minima without reaching a sensible solution. Only the fit kinetic parameters of the prevalent (dominant) reaction using HLPR model were finally presented in Extended Data Fig. 8g. For assessment,  $K_D$  values calculated from obtained kinetic parameters ( $k_d/k_a$ ) were compared with  $K_D$  values directly obtained using the LSS model (Extended Data Fig. 8g).

**Single molecule flow stretching assays: DNA substrate constructs.** Bacteriophage  $\lambda$  DNA was modified by ligating a biotinylated fork on one end and a digoxigenin moiety at the other end as described previously<sup>14</sup>. This ligated product was digested with either EcoRI or ApaI to generate 3.6- and 10.1-kb fragments from the forked and digoxigenin ends, respectively. An oligonucleotide sequence containing a single copy of wild type or variants of the TerB site was ligated to the digested ends of the 3.6- and 10-kb fragments as described previously<sup>34</sup> to generate DNA constructs with variant TerB sites that are listed in Table 1 and Extended Data Fig. 1b.

**Force calibration.** A force extension curve was constructed by measuring the length of individual 13.7 kb DNA molecules and calculating the hydrodynamic drag force at different flow rates using the equipartition theorem equation as described previously<sup>34,35</sup>. The force extension curve was fit using Worm-like chain model<sup>34,36</sup>. The fluctuation in the laminar flow causes an error in estimating the force and consequently the length of individual DNA molecules, which results in an error in estimating the location of the TerB site relative to the fork. At the applied stretching force of 2.6 pN in our experiments, the error in estimating the force, derived from the standard deviation among seven DNA molecules in the same field of view, results in an error of  $\pm \sim 85$  bp in estimating the position of the TerB site at 3.6 kb from the site of fork assembly<sup>34</sup>. Consequently, we treated any replication event ending between 3.5 and 3.7 kb as being stopped at the TerB site.

**Single-molecule leading-strand synthesis assay.** The leading strand DNA synthesis and data analysis were performed as described previously<sup>13,14</sup> with the variation of adding Tus to the reaction. Briefly, Tus was first introduced under continuous flow at 80 nM in buffer containing 30 mM Tris–HCl pH 7.6, 50 mM NaCl, 0.5 mM EDTA, 5 mM dithiothreitol and 10 mM MgCl<sub>2</sub> for 30 min to ensure the binding of Tus to TerB. The excess DNA-unbound Tus was removed

by washing with 15 times flow cell volume with replication buffer containing 50 mM HEPES-KOH pH 7.9, 80 mM KCl, 12 mM Mg(OAc)<sub>2</sub>, 2 mM MgCl<sub>2</sub>, 5 mM dithiothreitol and 0.1 mg ml<sup>-1</sup> BSA. Tus was then reintroduced with the replication proteins under continuous flow in the replication buffer supplemented with 760 μM of each dNTP, 1 mM ATP and proteins as follows: 80 nM Tus, 30 nM τ<sub>3</sub>δδ'χψ, 30 nM DnaB<sub>6</sub>(DnaC)<sub>6</sub> helicase-loader complex, 30 nM β<sub>2</sub> clamp, 60 nM αεθ core Pol III, and fork restart proteins, 20 nM PriA, 40 nM PriB and 480 nM DnaT. Experiments were carried out at 32 °C.

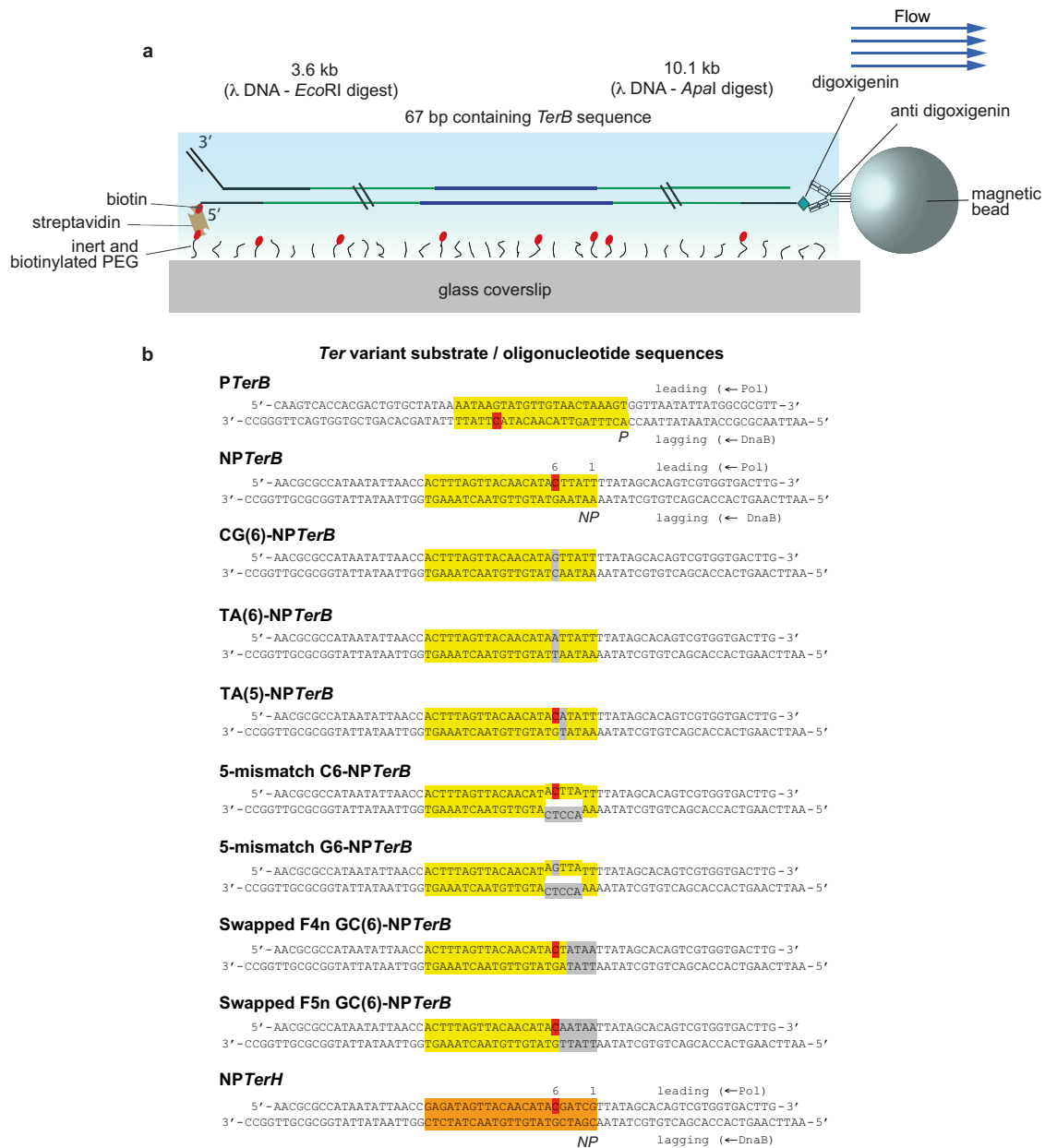
For data analysis, the picked particles were first corrected for their Brownian motion using unreplicated tethered DNA molecules. Pausing of DNA synthesis was considered when the amplitude fluctuations of a minimum of six data points (acquisition rate was 2 Hz) was less than three times the standard deviation of the noise. Bead displacement was converted into numbers of nucleotides synthesized using the known length difference between ss- and dsDNA in λ DNA<sup>35</sup> (3.76 bases per nm at our applied stretching force of 2.6 pN). Total experimental time was 30 min. In the study of the effect of Tus concentration on Tus-*TerB* polar arrest activity (Fig. 1f), Tus was first pre-incubated with the DNA at 80 nM and excess Tus was washed out as described above for our standard experimental condition. This was followed by the introduction of either 20 or 80 nM of Tus with the replication proteins. Tus(H144A) was used at concentration of 80 nM while Tus(R198A) was used at 250 nM throughout the reaction. Multiplexed single-molecule experimental results were derived from three or four technical replicates for each experimental condition.

The portion of leading strand synthesis trajectories that randomly terminated before reaching the position of the *Ter* site at 3.6 ± 0.1 kb (Extended Data Fig. 3a) were excluded from analysis. Those that reached the *Ter* site were separated into three categories: (1) those that continued unimpeded through the *Ter* site ('bypass'), (2) those that were 'permanently' arrested for all of the period of

observation (9.7 ± 1.0 min; 'stop'), and (3) those that paused (for ≥ 3 s, see above) and then resumed ('restart').

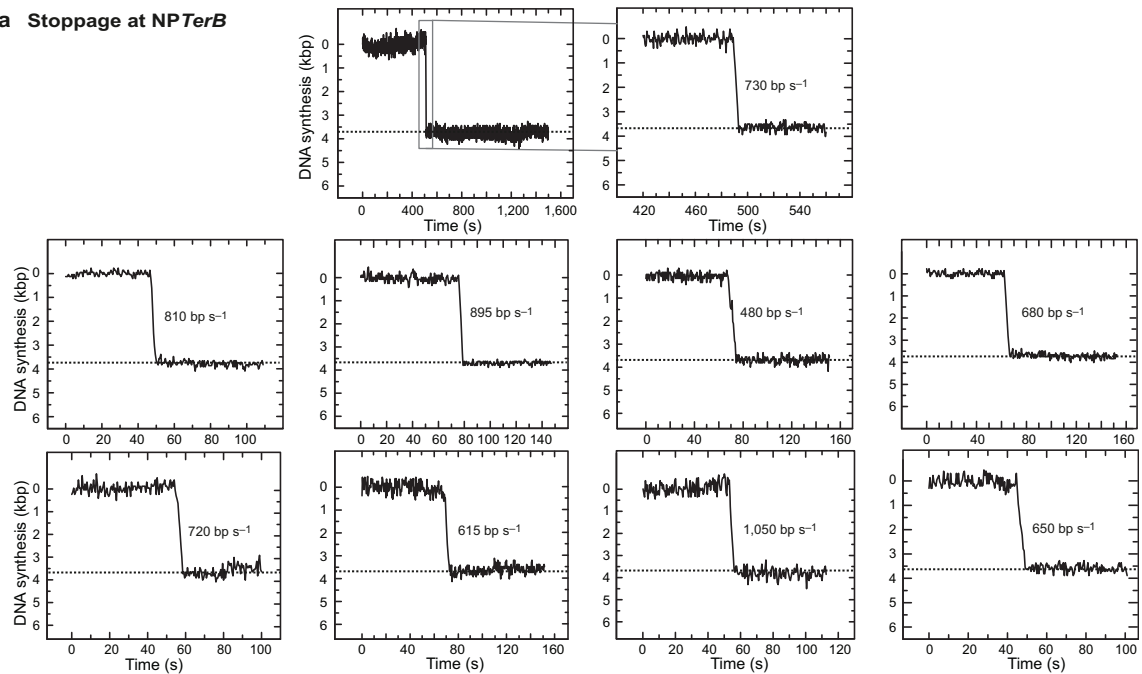
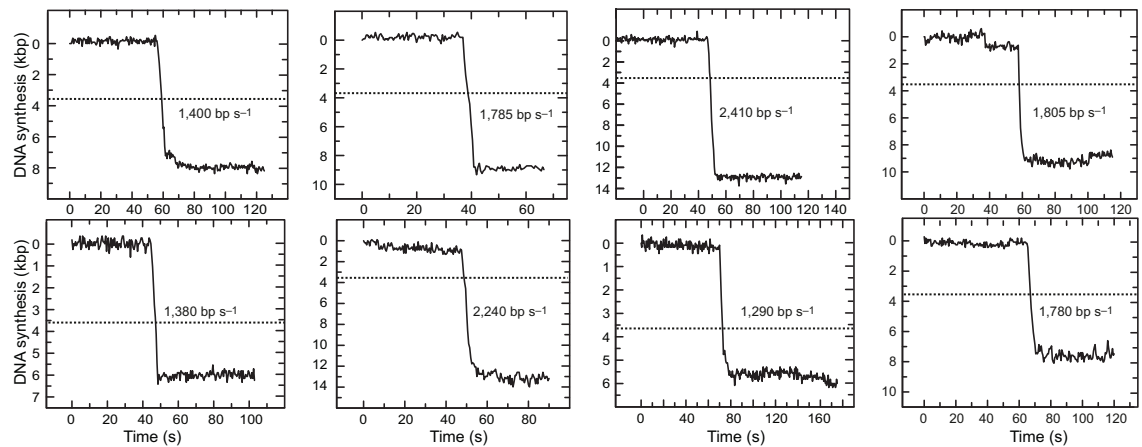
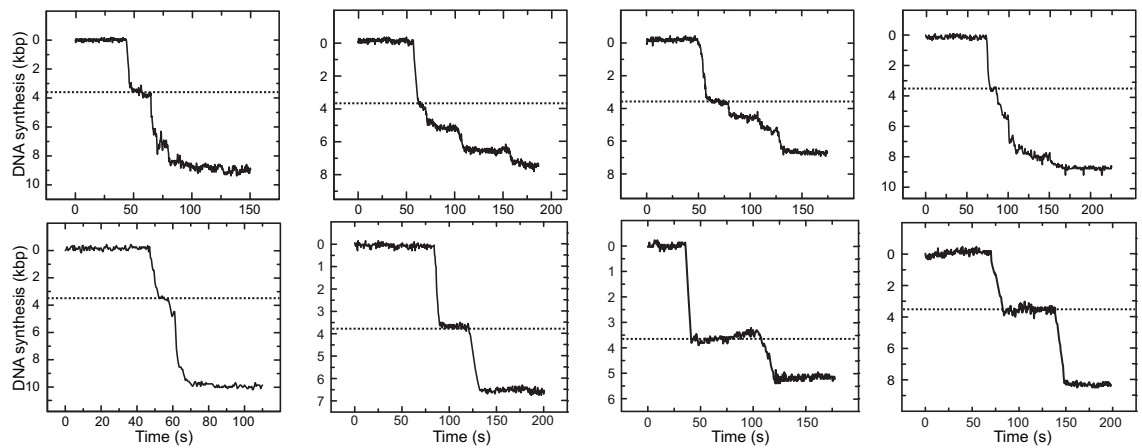
25. Oakley, A. J. *et al.* Flexibility revealed by the 1.85 Å crystal structure of the β sliding-clamp subunit of *Escherichia coli* DNA polymerase III. *Acta Crystallogr. D* **59**, 1192–1199 (2003).
26. Mariani, K. J. φX174-type primosomal proteins: purification and assay. *Methods Enzymol.* **262**, 507–521 (1995).
27. McPhillips, T. M. *et al.* Blu-Ice and the Distributed Control System: software for data acquisition and instrument control at macromolecular crystallography beamlines. *J. Synchrotron Radiat.* **9**, 401–406 (2002).
28. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
29. Vagin, A. & Teplyakov, A. MOLREP: an automated program for molecular replacement. *J. Appl. Crystallogr.* **30**, 1022–1025 (1997).
30. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D* **53**, 240–255 (1997).
31. Winn, M. D., Isupov, M. N. & Murshudov, G. N. Use of TLS parameters to model anisotropic displacements in macromolecular refinement. *Acta Crystallogr. D* **57**, 122–133 (2001).
32. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
33. The PyMOL Molecular Graphics System, Version 1.5.0.4, Schrödinger, LLC.
34. Pandey, M. *et al.* Two mechanisms coordinate replication termination by the *Escherichia coli* Tus-*Ter* complex. *Nucleic Acids Res.* **43**, 5924–5935 (2015).
35. van Oijen, A. M. *et al.* Single-molecule kinetics of lambda exonuclease reveal base dependence and dynamic disorder. *Science* **301**, 1235–1238 (2003).
36. Bustamante, C., Marko, J. F., Siggia, E. D. & Smith, S. Entropic elasticity of lambda-phage DNA. *Science* **265**, 1599–1600 (1994).





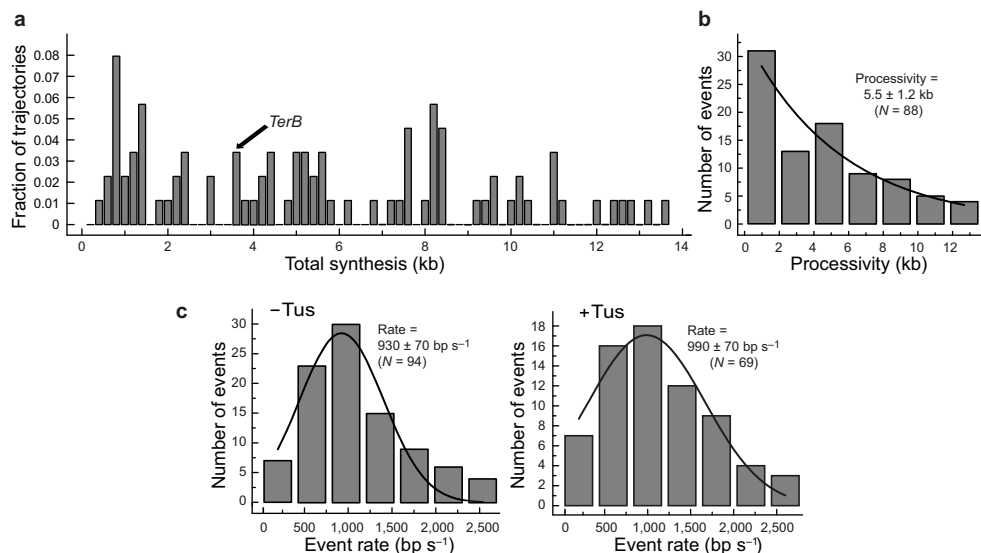
**Extended Data Figure 1 | Setup for leading-strand replication assays.** **a**, A schematic representation of the 13.7 kb DNA substrate construct. The substrate contains a biotinylated fork at one end to attach it to the streptavidin-coated glass coverslip and a digoxigenin moiety at the other end to attach it to a 2.8  $\mu$ m diameter anti-digoxigenin-coated paramagnetic bead. A single insert of *TerB* site is located at 3.6 kb from the biotinylated fork. **b**, Oligonucleotides used to assemble wild-type and variants of *TerB* substrates for their ligation to the 3.6 kb EcoRI and 10.1 kb ApaI  $\lambda$  DNA fragments<sup>34</sup>. Native *TerB* residues are

highlighted in yellow except C6 that is in red. Non-native (modified) residues in *TerB* are highlighted in grey. Native *TerH* residues are highlighted in orange. Leading and lagging DNA strands as well as permissive (P) and non-permissive (NP) faces of *Ter* when bound to Tus are denoted. Directionality of translocation of DnaB that encircles the lagging strand as it unwinds dsDNA during leading strand DNA synthesis by Pol III holoenzyme is denoted by arrows.

**a Stoppage at NP $\text{TerB}$** **b Bypass at NP $\text{TerB}$** **c Fate of forks at CG(6)-NP $\text{TerB}$** 

**Extended Data Figure 2 | Examples of trajectories for leading-strand synthesis upon encountering Tus bound to non-permissive *Ter* sites.** The location of the *TerB* site at  $3.6 \pm 0.1$  kb is indicated by the dashed lines. The rates of leading strand synthesis were calculated by fitting the slopes of the trajectories by linear regression using a least-squares approach. The replisomes displayed heterogeneity in rates of DNA synthesis. **a**, Trajectories where forks stopped at the NP*TerB* site. The average stoppage time captured within our acquisition time was  $9.7 \pm 1$  min (uncertainty is the standard error) as illustrated for the top trajectory. **b**, Trajectories where forks displaced Tus and

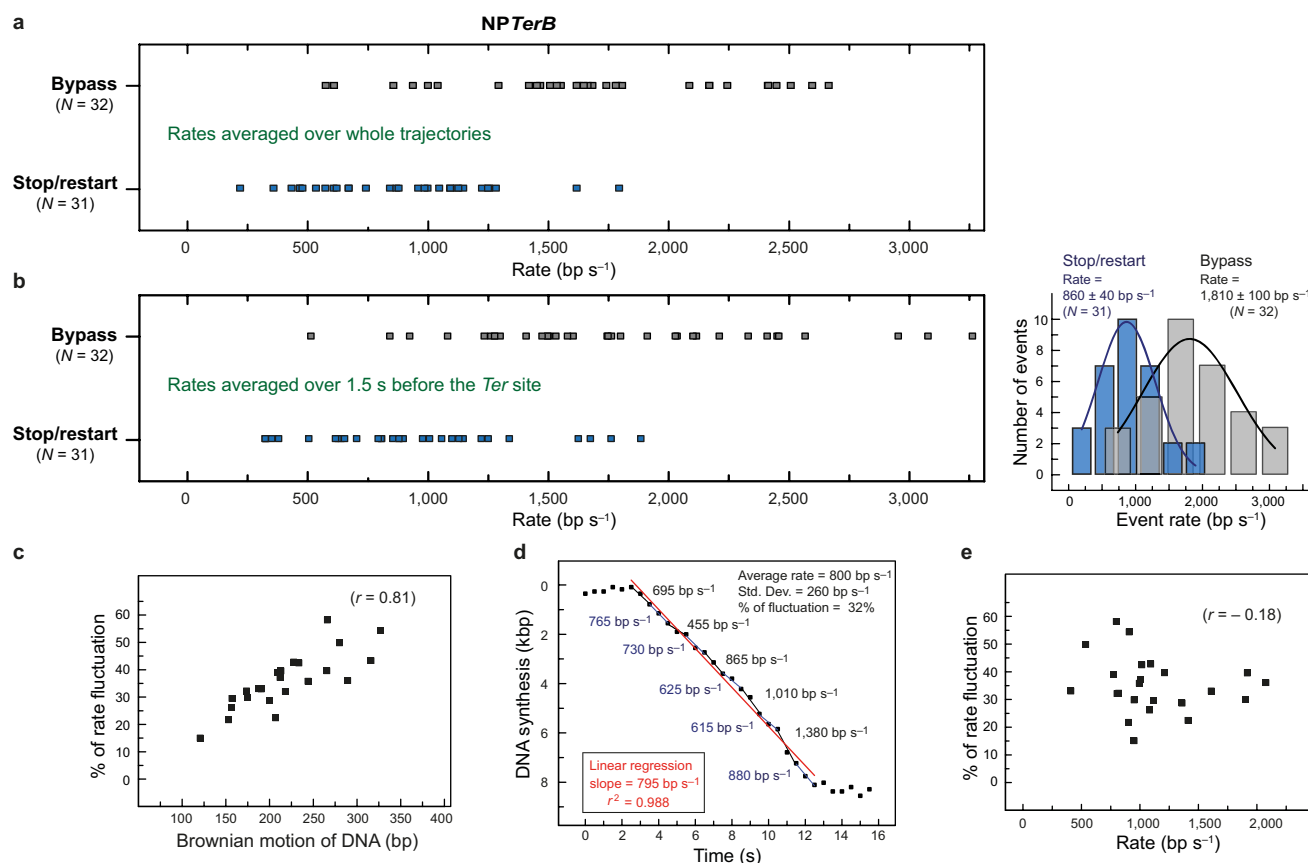
bypassed the NP*TerB* site without displaying any transient stoppage. **c**, Fate of the replication fork upon encountering CG(6)-NP*TerB*. Examples of trajectories for leading-strand synthesis upon encountering Tus bound to CG(6)-NP*TerB* showing transient stoppage at CG(6)-NP*TerB*, followed by resumption of DNA synthesis; 56% of the restarted events displayed DNA synthesis with disrupted behaviour (top row) while 44% showed normal behaviour (bottom row). We attributed the disrupted restart of DNA synthesis in some of the trajectories to the replisome losing some components other than DnaB during stalling.



**Extended Data Figure 3 | Effect of *TerB* site alone and nonspecifically DNA-bound Tus on DNA synthesis.** **a**, Probability of termination of DNA synthesis at 0.2 kb intervals (spatial resolution of the assay) along the 13.7 kb NP*TerB* in the absence of Tus, showing stops at *TerB* (3.5–3.7 kb, denoted by black arrow) occur randomly with a 3% probability when all events were considered, in contrast to 5% when only events that reached *TerB* ( $\geq 3.5$  kb) were taken into account. **b**, Processivity of DNA synthesis on the NP*TerB* substrate in the absence of Tus. The processivity distribution is fit with an exponential decay ( $N = 88$ ) and uncertainty corresponds to the standard error, illustrating the

random stoppage behaviour of the replisome during synthesis. **c**, Rate of leading strand synthesis using the 13.7 kb force-calibrated DNA construct (NP*TerB* in this case) in the absence (left panel;  $N = 94$ ) or presence of Tus (right panel;  $N = 69$ ). The rate distributions were fit with a Gaussian distribution. The fit lines are shown and the uncertainties correspond to the standard error. The rate agrees with our previously reported rate using force-calibrated  $\lambda$  DNA constructs<sup>14</sup>, demonstrating the accurate force calibration of the 13.7 kb substrate.



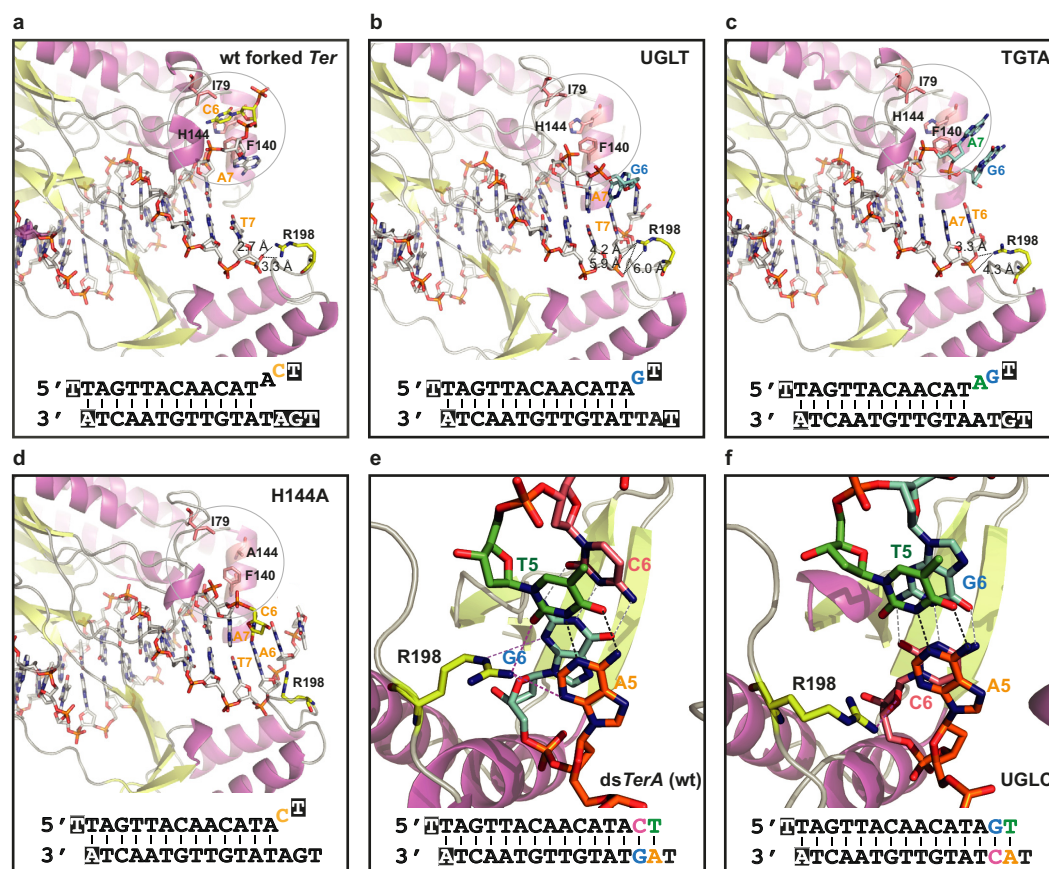


**Extended Data Figure 4 | Linear fitting of the rate of leading-strand synthesis is appropriate for deriving the correlation between rate of DNA synthesis and stalling activity at the NP*TerB* site.** **a**, Rate dependence of fork arrest at NP*TerB*. A scatter plot of forks that stopped ( $N = 31$ ) or bypassed ( $N = 32$ ) Tus bound to NP*TerB*; rates were calculated by fitting the DNA shortening phase of the entire trajectory in cases of events that bypassed and up to the stoppage point in events that stopped/restarted (histograms are shown in Fig. 1g). A significant correlation between fork progression rate and fork bypass at NP*TerB* is observed using a one sided Pearson's correlation test at the 0.05 level of significance (the calculated correlation coefficient ( $r$ ) was 0.62). The Pearson's correlation coefficient was calculated using the equation

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{[\sum_{i=1}^n (x_i - \bar{x})^2][\sum_{i=1}^n (y_i - \bar{y})^2]}}$$

**b**, Scatter plot (left) and rate distributions (right) of leading-strand synthesis for events that bypassed (grey bars) ( $N = 32$ ) or stopped/restarted (blue bars) ( $N = 31$ ) at NP*TerB* when the rate was estimated from fitting the slope of the three data points before the *TerB* site (acquisition time is 0.5 s per data point). The rates were fit with a Gaussian distribution and uncertainty corresponds to the standard error. The calculated average rates for events that bypassed or stopped/restarted at NP*TerB* are similar to those calculated when the rates were fit using the DNA shortening phase of the entire trajectory in cases of events that bypassed and up to the stoppage point in events that stopped/restarted (shown in Fig. 1g in the main text), underscoring the suitability of linear fitting of the rate.

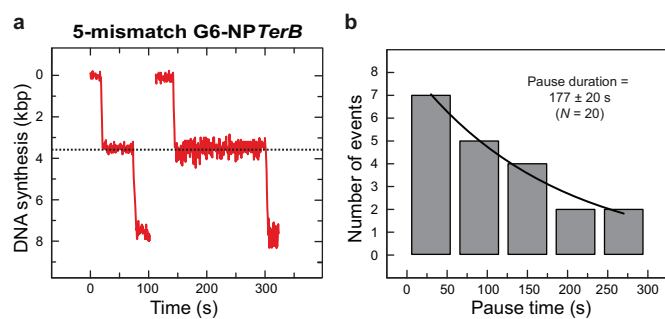
Furthermore  $r^2$  from linear regression fits was  $0.95 \pm 0.05$ . **c**, The correlation between apparent fluctuation in rate of DNA synthesis within individual DNA molecules and their corresponding Brownian motion ( $N = 23$ ). **d**, The individual trajectories displayed apparent fluctuation in rate of DNA synthesis as illustrated in a representative trajectory where we zoomed in at the DNA shortening phase and fit the rate linearly to intervals of three consecutive data points. The percentage of apparent fluctuation in rate of DNA synthesis within individual DNA molecules was calculated by dividing the standard deviation of the average of interval rates over the average rate. The standard deviation of the average of Brownian motion of each individual DNA molecule was calculated from the fluctuation of the DNA before and after being replicated. The percentage of apparent fluctuation in rate of individual DNA molecules displayed a strong positive correlation with their corresponding Brownian motion when analysed by two-sided Pearson's correlation test at the 0.05 level of significance ( $r = 0.81$ , panel c). **e**, The correlation between the percentage of apparent fluctuation in rate and the average rate of individual molecules. The percentage of apparent fluctuation in rate of individual molecules was calculated as described in d and for the same 23 replisomes. There was no correlation between the average rate of individual DNA molecules and their corresponding percentage of apparent fluctuation in rate; the Pearson's correlation coefficient was  $-0.18$ . The results from c–e demonstrate that one strong factor behind the apparent fluctuation in rates within our individual 13.7 kb molecules under our spatial and temporal resolution is the Brownian motion of the DNA and that this apparent fluctuation in rate does not bias the estimates of speed of the replisomes.



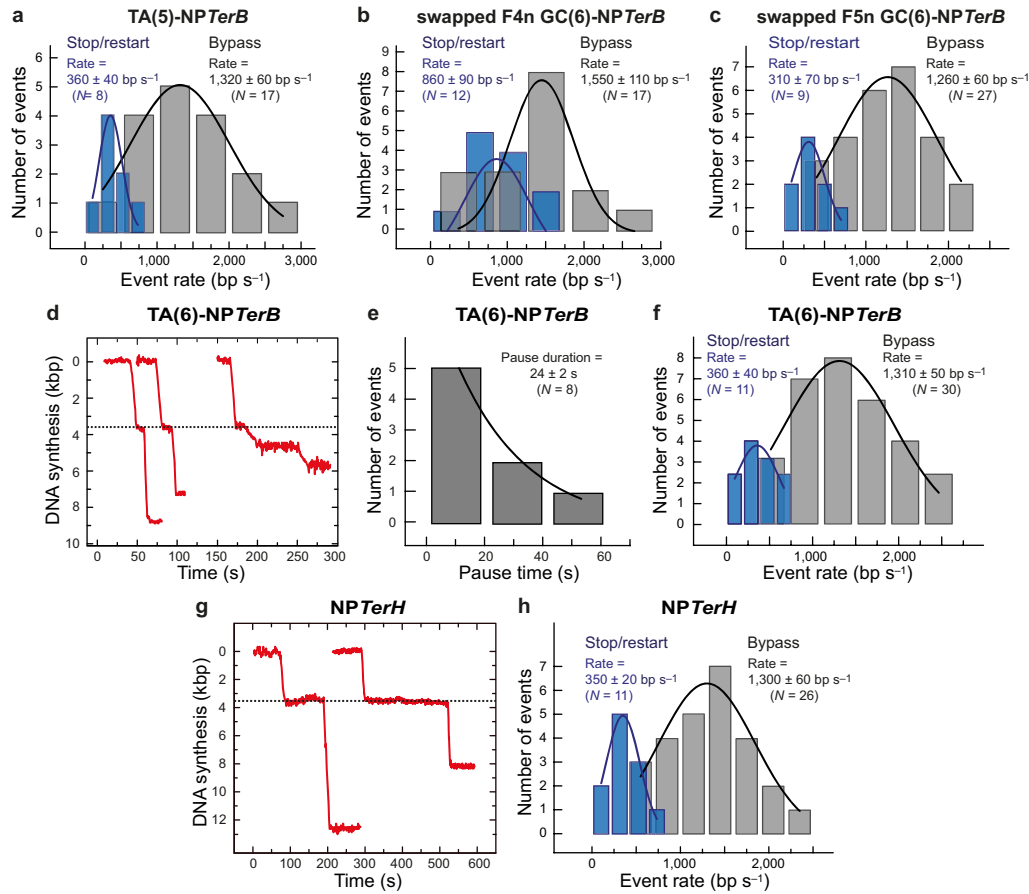
**Extended Data Figure 5 | Crystal structures of Tus complexes with *Ter* oligonucleotides.** The sequences of oligonucleotides used for each complex are shown at the bottom of each panel; nucleotides for which electron density could not be interpreted are highlighted. **a–d**, Complexes of Tus proteins with forked *Ter* sites. The C6-binding pocket is shown in the circle, with key residues Ile79, Phe140 and His144 in the binding pocket, and Arg198 shown in stick form. **a**, The wild-type Tus–*Ter* lock (PDB code: 2I06), with C6 located in the binding pocket, and the TA(7) base pair melted. Arg198 is positioned to interact with the 5′-phosphate of T7. **b**, Complex of wild-type Tus with a forked oligonucleotide that has C6 substituted by a mispaired G (UGLT: upper G, lower T; PDB code: 4XR0); G6 does not occupy the pocket nor does it make any new specific interactions with Tus, and Arg198 no longer interacts with the 5′-phosphate of T7. **c**, Further extension of the mismatched region in **b** to include A7 (TGTA: mispaired TGTA on the lower strand; PDB code: 4XR1) does not enable G6 to occupy the C6-binding pocket or form any new specific interactions. **d**, Tus(H144A) in complex with the normal Tus–*Ter* lock

oligonucleotide (PDB code: 4XR2), showing the mispaired C6 does not occupy the cytosine-binding pocket or form any new interactions with Tus.

**e, f**, Potential interactions of Arg198 in crystal structures of Tus complexes with fully base-paired *Ter* oligonucleotides. Only nucleotides in base pairs 5 and 6 are shown, and they are colour-coded to match the stick representations of them in the figures. Arg198 is shown in yellow stick representation. **e**, Structure of the wild-type Tus–*TerA* (GC(6)) complex (PDB code: 2I05). Arg198 is positioned potentially to make H-bonding interactions with the A5, G6 and T5 bases and the deoxyribose ring oxygen of G6, as well as electrostatic interactions with the 5′-phosphate of A5, as suggested previously<sup>11</sup> and demonstrated by molecular dynamics simulations (A.J.O., unpublished observations). **f**, Structure of the complex with a GC(6)-flipped version of the *TerA* oligonucleotide (UGLC: upper G, lower C; PDB code: 4XR3) showing an alternate major conformation of the Arg198 side-chain that has lost all base-specific interactions; only the interaction with the sugar ring oxygen of the substituted C6 is maintained.



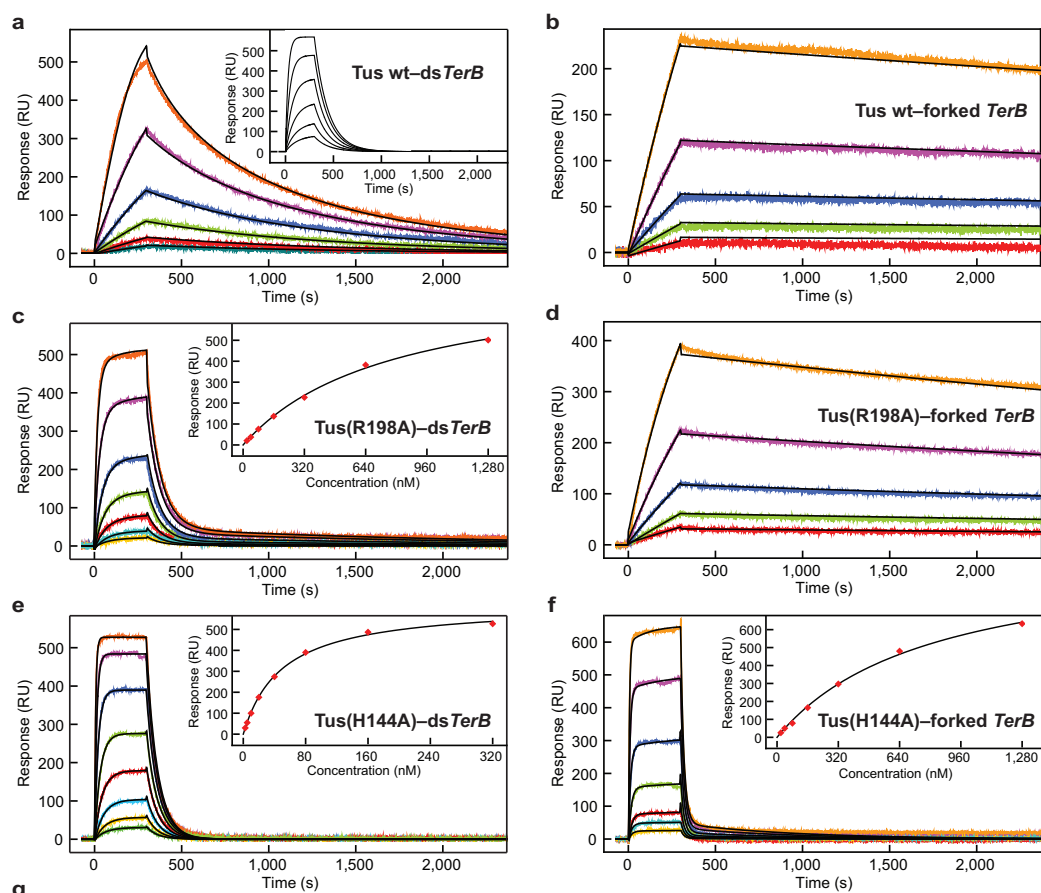
**Extended Data Figure 6 | Fate of the replication fork upon encountering 5-mismatch G(6)-NP TerB.** **a**, Examples of trajectories of replication forks that transiently stopped at Tus bound to the bubble template with C6 switched to G6 (5-mismatch G6-NP TerB). **b**, The distribution of the pause durations fit with a single exponential decay. The fit line is shown in black and the uncertainty corresponds to the standard error ( $N = 20$ ).



**Extended Data Figure 7 | Fate of the replication fork upon encountering NPterB sites with swapped sequences in the first five base pairs, TA(6)-NPTerB and NPterH.** Rate dependence of replication fork arrest at Tus bound to: **a**, TA(5)-NPTerB (N = 25); **b**, swapped F4n GC(6)-NPTerB (N = 29); **c**, swapped F5n GC(6)-NPTerB (N = 36). The rate distributions of leading-strand synthesis for events that bypassed (grey bars) or stopped/restarted (blue bars) at these sequences. **d**, Examples of trajectories of leading-strand synthesis that transiently stopped at Tus bound to TA(6)-NPTerB. 75% of the restarted events displayed DNA synthesis of normal behaviour (left traces) while 25% showed disrupted behaviour (right trace). **e**, The distribution

of the pause durations at TA(6)-NPTerB fit with a single exponential decay (N = 8). **f**, The rate distribution of events that bypassed (N = 30; grey bars) or stopped/restarted (N = 11; blue bars) at TA(6)-NPTerB. **g**, Examples of trajectories of leading-strand synthesis that transiently stopped at Tus bound to NPterH. The average pause duration was  $180 \pm 26$  s (N = 4). The uncertainty is the standard error. **h**, The rate distribution of leading-strand synthesis for events that bypassed (N = 26; grey bars) or stopped/restarted (N = 11; blue bars) at NPterH. The histograms in **a–c**, **e**, **f** and **h** were fit to Gaussian distributions, the fit lines are shown, and the uncertainties correspond to the standard error.

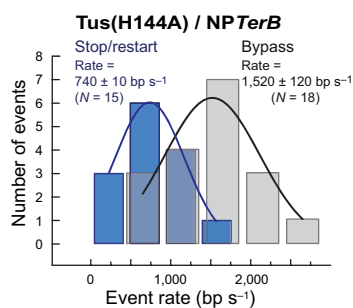




Panel	Interaction	[Tus] (nM)	$K_D$ (nM)	$k_a$ ( $M^{-1}s^{-1}$ )	$k_d$ ( $s^{-1}$ )	$t_{1/2}$ (s)
<b>a</b>	Tus–dsTerB	0.25–8	$1.86 \pm 0.08$	$(3.78 \pm 0.08) \times 10^6$	$(7.04 \pm 0.15) \times 10^{-3}$	98
<b>a (inset)</b>	Tus–dsTerB (simulated)	0.25–8	1.86	$3.78 \times 10^6$	$7.04 \times 10^{-3}$	98
<b>b</b>	Tus–forked TerB	0.25–4	$0.212 \pm 0.002$	$(2.89 \pm 0.00) \times 10^5$	$(6.12 \pm 0.07) \times 10^{-5}$	11,320
<b>c</b>	Tus(R198A)–dsTerB	20–1,280	$441 \pm 10$	$(2.99 \pm 0.00) \times 10^4$	$(1.32 \pm 0.03) \times 10^{-2}$	53
<b>c (inset)</b>	Tus(R198A)–dsTerB	20–1,280	$780 \pm 60$	—	—	—
<b>d</b>	Tus(R198A)–forked TerB	2.5–40	$1.80 \pm 0.01$	$(5.52 \pm 0.00) \times 10^4$	$(9.95 \pm 0.03) \times 10^{-5}$	6,970
<b>e</b>	Tus(H144A)–dsTerB	2.5–320	$39.3 \pm 1.0$	$(3.92 \pm 0.02) \times 10^5$	$(1.54 \pm 0.03) \times 10^{-2}$	45
<b>e (inset)</b>	Tus(H144A)–dsTerB	2.5–320	$49.5 \pm 2.3$	—	—	—
<b>f</b>	Tus(H144A)–forked TerB	20–1,280	$598 \pm 8$	$(9.09 \pm 0.07) \times 10^4$	$(5.44 \pm 0.03) \times 10^{-2}$	13
<b>f (inset)</b>	Tus(H144A)–forked TerB	20–1,280	$805 \pm 70$	—	—	—

**Extended Data Figure 8 | SPR assessment of Tus–*TerB* interactions: whereas Tus and Tus(R198A) are capable of forming a lock, Tus(H144A) is not.** ProteOn sensorgrams show association and dissociation phases of Tus–*TerB* interactions at ranges of Tus concentrations (as specified in **g**) of serially-diluted samples of Tus proteins. Curves, shown in colours, were fit simultaneously (black curves) to various binding models (see Methods). **a**, Wild-type Tus and ds*TerB*. Considering that the  $k_a > 1 \times 10^6 \text{ M}^{-1} \text{ s}^{-1}$  suggests significant mass transport limitations, the LMT model was used to fit the data with  $R_{\text{max}}$  constrained to 700 RU. The derived kinetic parameters were used to simulate sensorgrams devoid of mass transfer limitation using the L model (inset). **b**, Wild-type Tus–forked *TerB* interaction;  $R_{\text{max}}$  was constrained to 775 RU. The fit  $k_d$  is in good agreement with the value of  $(5.20 \pm 0.00) \times 10^{-5} \text{ s}^{-1}$  obtained from an independent experiment where dissociation was monitored over 50,000 s (not shown). **c**, Tus(R198A)–ds*TerB* interaction. Binding kinetics parameters were obtained using the HLP model. The sum of fit  $R_{\text{max}1}$  ( $543 \pm 9$ ) and  $R_{\text{max}2}$  ( $54 \pm 5$  RU) values were in reasonable agreement with the expected value of  $\sim 700$  RU. Only the relevant  $k_a$  and  $k_d$  values of the predominant (based on  $R_{\text{max}1}$ ) interaction are presented in **g**. For assessment of the fitting procedure, responses at equilibrium were fit using the L model (inset). The derived  $K_D$  was within the factor of two of the calculated  $K_D$  obtained from kinetic parameters ( $k_d/k_a$ ). The  $R_{\text{max}}$  value of  $816 \pm 32$  RU was slightly higher than theoretical (700 RUs), probably owing to some non-specific binding in the high range of Tus concentration. **d**, Tus(R198A)–forked *TerB* interaction. The L model was used to fit the

data with  $R_{\text{max}}$  constrained to 775 RU. The fit  $k_d$  was within a factor of two of the value,  $(5.70 \pm 0.00) \times 10^{-5} \text{ s}^{-1}$ , derived from an independent experiment where dissociation was monitored over 50,000 s (not shown). **e**, Tus(H144A)–ds*TerB* interaction. Binding kinetic parameters were obtained using the HLP model. The sum of fit  $R_{\text{max}1}$  ( $537 \pm 1$ ) and  $R_{\text{max}2}$  ( $31 \pm 0$  RU) values were in reasonable agreement with the expected value of  $\sim 700$  RU. Only the relevant  $k_a$  and  $k_d$  values of the predominant interaction ( $R_{\text{max}1}$ ) are presented in **g**. For assessment of the fitting procedure, responses at equilibrium were fit using the L model (inset). The derived  $K_D$  was within a factor of 1.5 of  $K_D$  obtained from the kinetic parameters. In addition, the fit  $R_{\text{max}}$  value of  $621 \pm 10$  RU compares reasonably to the expected value of  $\sim 700$  RU. **f**, Tus(H144A)–forked *TerB* interaction. Binding kinetics parameters were obtained using the HLP model. The sum of fit  $R_{\text{max}1}$  ( $879 \pm 4$ ) and  $R_{\text{max}2}$  ( $65 \pm 1$ ) values were somewhat high compared to the expected value of  $\sim 775$  RU. Only the relevant  $k_a$  and  $k_d$  values of the predominant reaction are presented in **g**. Responses at equilibrium were fit using the L model (inset). Derived  $K_D$  was within the factor of 2 of the calculated  $K_D$  obtained from ( $k_d/k_a$ ). In addition, fit  $R_{\text{max}}$  value of  $1,040 \pm 50$  RU was slightly higher than theoretical. **g**, Summary of binding parameters for Tus–*Ter* interactions. All uncertainties are standard errors in parameters from fitting of complete data sets to appropriate binding models as described in the Methods. Data are representative of those from two technical replicates using different instruments (BiaCore T200 and ProteOn XPR-36).



**Extended Data Figure 9 | Fate of the replication fork upon encountering Tus(H144A) bound to NP $\text{TerB}$ .** Rate dependence of fork arrest. The rate distribution of leading-strand synthesis for events that bypassed ( $N = 18$ ; grey bars) or stopped/restarted ( $N = 15$ ; blue bars) at NP $\text{TerB}$  fit with Gaussian distributions. The fit lines are shown and the uncertainties correspond to the standard error.

Extended Data Table 1 | Data collection and refinement statistics for Tus–Ter complexes

	wt Tus / forked <i>Ter</i> -UGLT [C6 to G mutant]	wt Tus / forked <i>Ter</i> -TGTA [C6 to G, T7 to A mutant]	Tus(H144A) / wt forked <i>Ter</i>	wt Tus / ds <i>Ter</i> -UGLC [GC(6) to CG(6) mutant]
<b>Data collection</b>				
Space group	<i>P</i> 4 <sub>1</sub> 2 <sub>1</sub> 2	<i>P</i> 4 <sub>1</sub> 2 <sub>1</sub> 2	<i>P</i> 4 <sub>1</sub> 2 <sub>1</sub> 2	<i>P</i> 4 <sub>1</sub> 2 <sub>1</sub> 2
Cell dimensions				
<i>a</i> , <i>b</i> , <i>c</i> (Å)	64.5, 64.5, 248.3	64.8, 64.8, 246.7	64.5, 64.5, 250.9	64.1, 64.1, 249.3
$\alpha$ , $\beta$ , $\gamma$ (°)	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90
Resolution (Å)	75–2.80 (2.90–2.80)	75–2.40 (2.50–2.40)	75–2.35 (2.43–2.35)	75–2.70 (2.80–2.70)
<i>R</i> <sub>sym</sub>	11.1 (74.0)	12.0 (71.1)	8.0 (85.6)	10.6 (84.1)
<i>I</i> / $\sigma$ <i>I</i>	39.5 (6.6)	24.0 (3.3)	26.7 (2.0)	27.2 (3.6)
Completeness (%)	99.8 (99.6)	99.9 (100)	99.6 (100)	99.8 (100)
Redundancy	22.1 (18.9)	9.7 (10.8)	6.6 (7.5)	13.6 (13.4)
<b>Refinement</b>				
Resolution (Å)	45–2.80 (2.86–2.80)	63–2.40 (2.46–2.40)	62–2.35 (2.41–2.35)	62–2.70 (2.77–2.70)
No. reflections	12,822 (901)	20,288 (1,521)	21,819 (1,658)	14,304 (1,086)
<i>R</i> <sub>work</sub> / <i>R</i> <sub>free</sub>	21.7 / 29.6	19.8 / 26.9	22.0 / 26.6	20.7 / 28.4
No. atoms				
Protein	2,504	2,530	2,509	2,503
Nucleic acid	577	574	592	612
Ligand/ion	13	14	14	12
Water	31	96	69	32
<i>B</i> -factors				
Protein	25.3	43.3	39.3	61.3
Nucleic acid	26.5	45.1	38.4	57.4
Ligand/ion	28.7	50.3	58.1	61.4
Water	35.7	57.2	48.1	61.3
R.m.s deviations				
Bond lengths (Å)	0.016	0.017	0.013	0.013
Bond angles (°)	2.16	2.13	1.71	1.97

A single crystal was used in each case.

Numbers in parentheses refer to the highest resolution bin.



# Integrator mediates the biogenesis of enhancer RNAs

Fan Lai<sup>1\*</sup>, Alessandro Gardini<sup>1\*</sup>, Anda Zhang<sup>1</sup> & Ramin Shiekhattar<sup>1</sup>

**Integrator is a multi-subunit complex stably associated with the carboxy-terminal domain (CTD) of RNA polymerase II (RNAPII)<sup>1</sup>. Integrator is endowed with a core catalytic RNA endonuclease activity, which is required for the 3'-end processing of non-polyadenylated, RNAPII-dependent, uridylate-rich, small nuclear RNA genes<sup>1</sup>. Here we examine the requirement of Integrator in the biogenesis of transcripts derived from distal regulatory elements (enhancers) involved in tissue- and temporal-specific regulation of gene expression in metazoans<sup>2–5</sup>. Integrator is recruited to enhancers and super-enhancers in a stimulus-dependent manner. Functional depletion of Integrator subunits diminishes the signal-dependent induction of enhancer RNAs (eRNAs) and abrogates stimulus-induced enhancer–promoter chromatin looping. Global nuclear run-on and RNAPII profiling reveals a role for Integrator in 3'-end cleavage of eRNA primary transcripts leading to transcriptional termination. In the absence of Integrator, eRNAs remain bound to RNAPII and their primary transcripts accumulate. Notably, the induction of eRNAs and gene expression responsiveness requires the catalytic activity of Integrator complex. We propose a role for Integrator in biogenesis of eRNAs and enhancer function in metazoans.**

To assess the role for Integrator in the biogenesis of eRNAs, we examined the signal-dependent recruitment of Integrator complex to enhancer sites. HeLa cells were starved of serum for 48 h, after which they were stimulated with epidermal growth factor (EGF) to induce immediate early genes (IEGs). We identified 2,029 enhancers based on their occupancy by RNAPII, CBP/p300 and containing acetylated histone H3 lysine 27 (H3K27ac) chromatin modification (see Methods). We found that while assessing steady-state levels of eRNAs provided a measure of EGF-induced eRNAs, we obtained a better read-out of eRNAs after sequencing of the chromatin-enriched RNA fractions (ChromRNA-seq)<sup>6</sup>. We focused on 91 enhancers that displayed EGF-induced eRNAs in the proximity of EGF-responsive genes following 20 min of induction (Extended Data Fig. 1, Supplementary Table 1 and see Methods). Notably, the chromatin surrounding these enhancers displayed the H3K27ac modification in starved cells, and following EGF stimulation there was a small increase in H3K27ac levels (Extended Data Fig. 1b). To assess the polyadenylation state of eRNAs, total RNA was enriched for polyadenylated and non-polyadenylated fractions and was subjected to high-throughput sequencing. Similar to previous reports, EGF-induced enhancers displayed bi-directional eRNAs that were predominantly not polyadenylated (Extended Data Fig. 2)<sup>5,7</sup>.

We next analysed Integrator occupancy at these enhancers by using antibodies against the INTS11 subunit of the Integrator complex before and after EGF stimulation. While these enhancers were occupied by a detectable amount of Integrator before EGF induction, addition of EGF resulted in a further recruitment of Integrator complex (Fig. 1a–c). RNAPII displayed a similar pattern of stimulus-dependent chromatin residence (Fig. 1d, e). The stimulus-dependent recruitment of Integrator at enhancers was further confirmed using two additional antibodies against INTS1 and INTS9 subunits of the Integrator complex (Extended Data Fig. 3a). These results demonstrated the

stimulus-dependent recruitment of the Integrator complex at EGF-responsive enhancers.

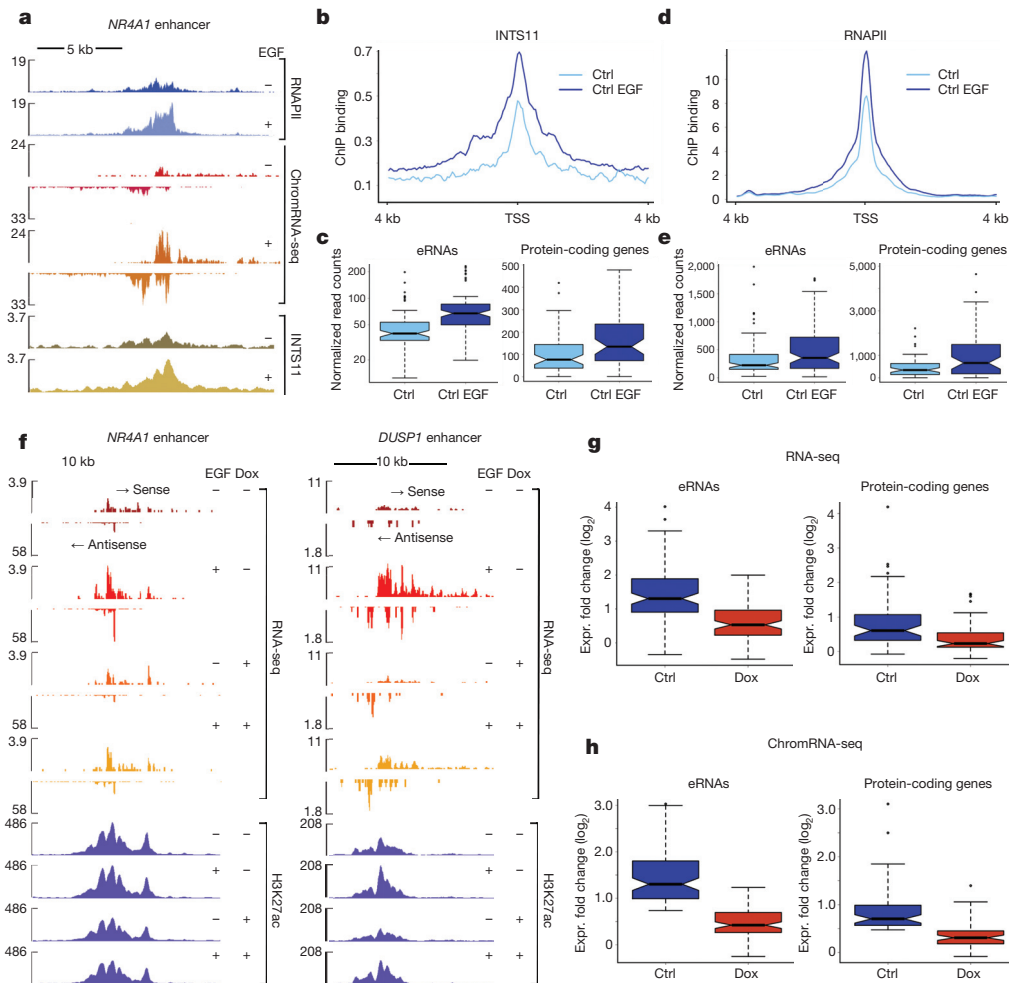
To examine the functional importance of Integrator at enhancers and its role in the biogenesis of eRNAs, we developed HeLa clones expressing doxycycline-inducible short hairpin RNAs (shRNAs) against INTS11 and INTS1 subunits of the Integrator complex (Extended Data Fig. 3b). Within the time course of these experiments the mature levels of small nuclear RNAs (snRNAs) were not perturbed (data not shown). Twenty minutes of EGF stimulation resulted in the induction of bi-directional eRNAs similar to previous reports (Fig. 1a, f and Extended Data Fig. 1c–h)<sup>5,8–11</sup>. Depletion of INTS11 diminished the eRNA induction after EGF stimulation (Fig. 1f; as shown at two enhancer loci; enhancers were named after their proximity to an EGF-responsive gene). The fold induction of eRNAs at all EGF-induced enhancers decreased significantly (Fig. 1g, h). We also observed a significant decrease in the transcriptional induction of EGF-responsive protein-coding genes in the proximity of these EGF-induced enhancers (Fig. 1g, h). Interestingly, there was a subtle increase (statistically not significant) in H3K27 acetylation at enhancers following EGF stimulation, which was reduced after Integrator depletion (Fig. 1f and Extended Data Fig. 3c).

To gain further insight into quantitative changes in eRNAs following depletion of Integrator, we depleted INTS11 or INTS1 and performed a time-course analysis of eRNA induction using specific primer sets for each strand. Depletion of either Integrator subunit diminished the EGF-induced increase in eRNA levels from both strands of the enhancers (Extended Data Fig. 4a, b). Analysis of regulatory landscape in the proximity of the EGF-responsive gene *ATF3* (activating transcription factor 3) revealed the presence of clusters of acetylated H3K27 and p300 binding sites similar to that described for super-enhancers<sup>12–14</sup> (Extended Data Fig. 4c). This region also displayed occupancy by RNAPII at multiple sites, and we could detect additional recruitment of RNAPII and Integrator to these sites following EGF stimulation (Extended Data Fig. 4c). Analysis of eRNA synthesis using strand-specific RNA-seq and real-time PCR (during a time-course experiment) demonstrated a requirement for Integrator in the induction of eRNAs at the super-enhancer sites after EGF stimulation (Extended Data Fig. 4d). Collectively, these results highlight a requirement for Integrator in stimulus-dependent induction of eRNAs from individual enhancers and enhancer clusters.

An important component of enhancer function is the formation of stimulus-dependent chromatin looping, allowing enhancer and promoter communication<sup>15–17</sup>. We measured chromatin looping between *NR4A1* and *DUSP1* enhancers and their respective promoters using chromosome conformation capture (3C) following stimulation with EGF (Fig. 2a). We observed a robust association between the enhancer and the promoter regions of *NR4A1* and *DUSP1* after EGF stimulation (Fig. 2b). Remarkably, depletion of Integrator abrogated the EGF-induced chromatin looping without any effect on non-stimulus-induced chromosomal interactions (Fig. 2b, c and Extended Data Fig. 5a, b). These results demonstrate that Integrator regulates

<sup>1</sup>University of Miami Miller School of Medicine, Sylvester Comprehensive Cancer Center, Department of Human Genetics, Biomedical Research Building, Room 719, 1501 NW 10th Avenue, Miami, Florida 33136, USA.

\*These authors contributed equally to this work.



**Figure 1 | Integrator mediates induction of eRNAs.** **a**, EGF induction of an enhancer in the vicinity of the *NR4A1* gene (see Extended Data Fig. 1i). RNAPII and INTS11 are recruited to the enhancer after 20 min of stimulation and eRNAs are transcribed bi-directionally from the locus (as revealed by deep sequencing of chromatin-associated RNA, ChromRNA-seq). The y axis represents the read counts normalized to sequencing depth. **b**, Average profile of Integrator recruitment to 91 EGF-responsive enhancers. TSS indicates transcription start site. The y axis shows the average of read density. **c**, Increased Integrator occupancy at enhancers and their corresponding protein-coding genes (mean density was calculated as follows: 6 kb surrounding the peak of RNAPII for eRNAs; from  $-0.5$  kb to  $+2.5$  kb for coding genes;  $P < 0.001$ ). Whiskers on the box plot indicate the variability in the datasets. **d**, Average profile of RNAPII upon EGF treatment at enhancers. **e**, Increased RNAPII

enhancer function as reflected by the physical association between enhancers and their respective promoters.

To gain an insight into the mechanism by which Integrator regulates enhancer function and eRNA biogenesis, we depleted Integrator and performed RNAPII profiling and global nuclear run-on followed by high-throughput sequencing (GRO-seq) after EGF induction. Notably, Integrator depletion resulted in the increase and spreading of GRO-seq reads throughout the body of eRNA transcripts at both enhancers and super-enhancers, which was mirrored by a concomitant increase and spreading of RNAPII localization (Fig. 3a, b). Indeed, the average profile of depth-normalized reads of 91 EGF-induced enhancers showed a significant accumulation of GRO-seq and RNAPII ChIP-seq reads (Extended Data Fig. 6a, b). Analysis of RNAPII travelling ratio, a measure of RNAPII productive elongation, revealed that in contrast to EGF-responsive protein coding genes, which experience a block in productive elongation after Integrator depletion<sup>18</sup>, there is increased RNAPII occupancy in the body of eRNA transcripts (Extended Data Fig. 6c, d). The accumulation of

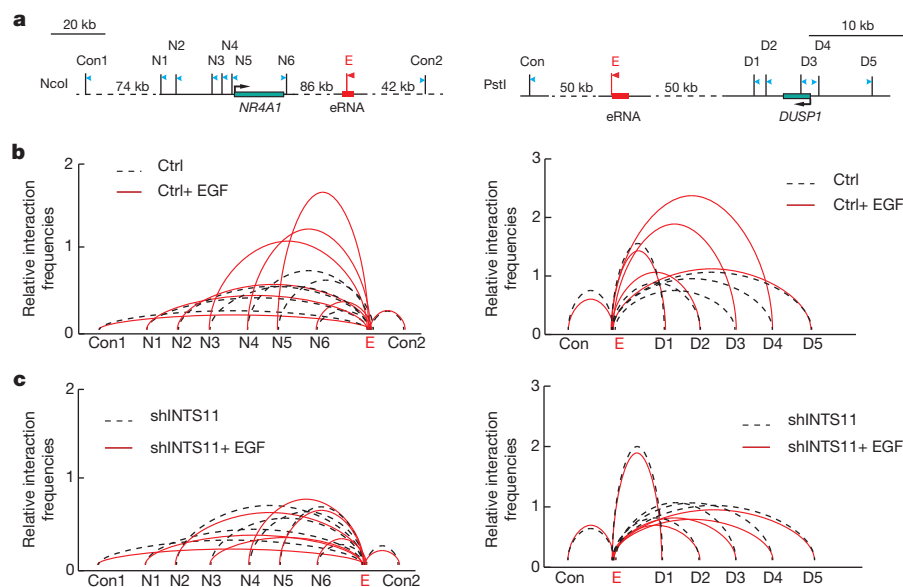
occupancy following EGF stimulation at enhancers and their corresponding protein-coding genes ( $P < 0.005$ ). **f**, Inducible knockdown of INTS11 (doxycycline (dox)) markedly reduces steady-state levels of eRNAs (as measured by total RNA-seq). Data were obtained using a tet-inducible shRNA system, stably transduced in HeLa cells. Acetylation of H3K27 is also shown. **g**, **h**, Average expression levels of 91 eRNAs and their neighbouring (<500 kb) 57 protein-coding genes indicate a significant impairment of activation. Box plots represent the expression fold change ( $\log_2$ ) before and after EGF treatment in normal conditions (ctrl) and upon depletion of Integrator (dox) ( $t$ -test,  $P < 0.0005$  for all panels). Fold change of RPKM (reads per kilobase of exon per million mapped reads) values was calculated from RNA-seq (**f**) and ChromRNA-seq (**g**) data.

RNAPII at eRNA loci after Integrator depletion occurred despite the decreased recruitment of super elongation complex (SEC) to enhancers (Extended Data Fig. 7a, b).

The increased RNAPII occupancy at eRNA loci suggests a block in 3'-end cleavage of primary eRNA transcripts, leading to a defect in termination. To quantitate such a 3'-end cleavage defect, we measured the accumulation of primary levels (or unprocessed levels) of eRNA transcripts after Integrator depletion using semi-quantitative PCR and real-time PCR. We observed a 3- to 10-fold accumulation of unprocessed eRNA transcripts concomitant with the reduction of the processed eRNA levels (Fig. 3c-e and Extended Data Fig. 8a). Previous experiments revealed that the loss of 3'-end cleavage by Integrator led to increased levels of polyadenylated U snRNA transcripts, which are normally not polyadenylated<sup>19</sup>. Indeed, analysis of the polyadenylated transcripts revealed a robust increase in polyadenylation of eRNAs in the absence of Integrator (Fig. 3f, g). These results attest to Integrator cleavage of the 3' end of eRNAs leading to a termination of transcription.

**Figure 2 | Integrator is required for enhancer-promoter interaction.**

**a**, Diagrams of *NR4A1* (left) and *DUSP1* (right) genomic regions with their respective enhancers (shown in red). The arrowheads depict the position of primers for detection of chromatin looping and the stick bars indicate enzyme digestion sites (named N1–6 and D1–5). E refers to the anchor primer at the enhancer sites; control sites are also indicated. **b**, Looping events between the promoter region of *NR4A1* and its enhancer were detected at N3, N4 and N5 sites after EGF induction (left). A similar interaction was also captured between sites D3 and D4 of *DUSP1* promoter and its downstream enhancer after EGF induction (right). **c**, Knockdown of Integrator abolished chromosomal looping events at both *NR4A1* and *DUSP1* sites. The interaction frequency between the anchoring points and the distal fragments were determined by real-time PCR and normalized to BAC templates. All sites were assayed in three independent experiments ( $P < 0.01$ , two-sided  $t$ -test). Control anchors are displayed in Extended Data Fig. 5.

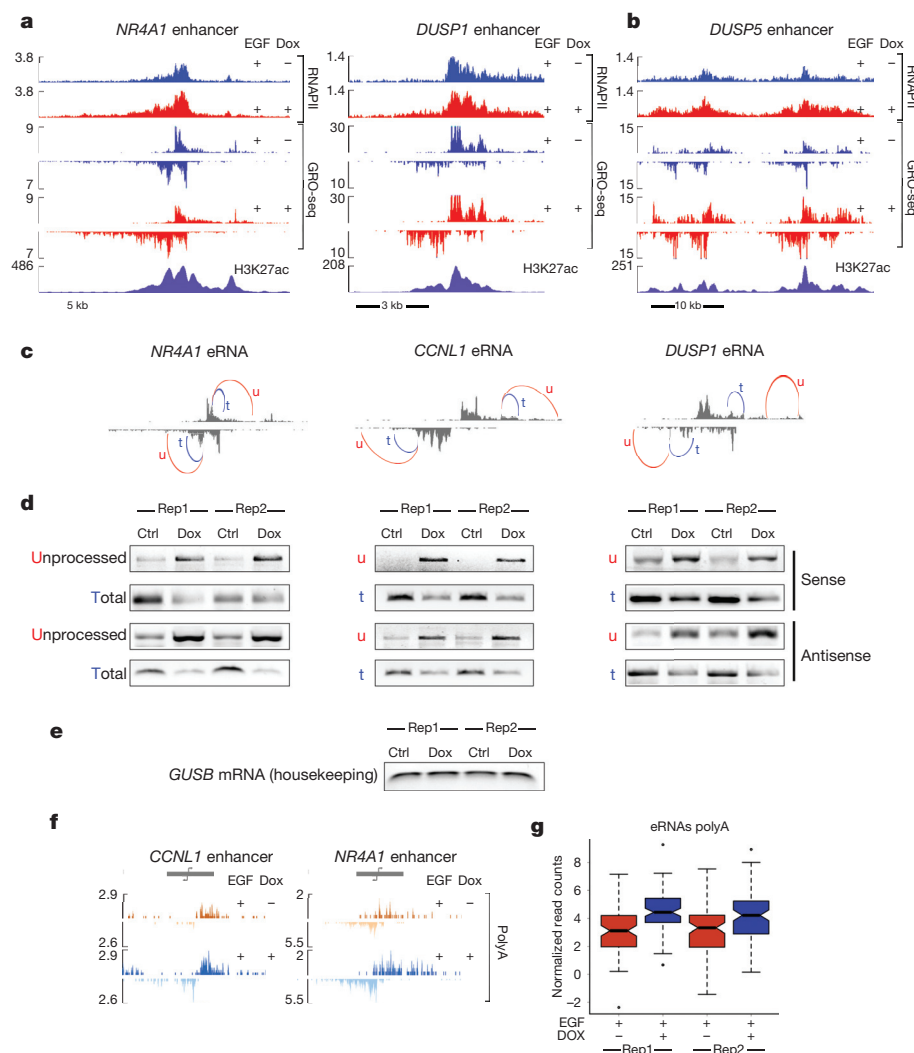


We surmised that such a termination defect might result in the inability of RNAPII to dissociate from the eRNAs, leading to accumulation of RNAPII–eRNA complexes and a consequent decrease in mature eRNA levels. We performed ultraviolet (UV) cross-linking followed by RNA

immunoprecipitation (UV-RIP) using antibodies against RNAPII to examine increased association of eRNAs with RNAPII after depletion of Integrator. Consistent with a role for Integrator in the processing of eRNAs, depletion of Integrator led to a profound increase in eRNA

**Figure 3 | Integrator has a role in termination of eRNAs.**

**a, b**, RNAPII dynamics was analysed by ChIP-seq and GRO-seq at the enhancer regions adjacent to *NR4A1* and *DUSP1* (**a**) and at the super-enhancer upstream of *DUSP5* (**b**). The y axis represents the read counts normalized to sequencing depth. **c**, 3'-end cleavage of eRNAs was examined with semi-quantitative PCR. Primer pairs were designed to amplify a portion of the enhancer transcript as detected in the control GRO-seq experiment (t, total) or a longer template further extending into the 3' of the enhancer region (u, unprocessed). **d**, PCR analysis was performed in two independent replicates, before (ctrl) and after (dox) depletion of INTS11 at three eRNAs (sense and antisense strand). **e**, The housekeeping gene *GUSB* was used as a cDNA loading control. **f**, Polyadenylation of eRNAs increases after depletion of Integrator at *DUSP1* and *CCNL1* enhancer loci. The polyadenylated fraction of RNA from whole-cell lysates was sequenced after EGF stimulation, before and after depletion of INTS11 (dox). **g**, Box plot shows significant increase in polyadenylated RNA reads ( $P < 0.001$ ) across the entire set of EGF responsive enhancers. Whiskers on the box plot indicate the variability in the datasets.



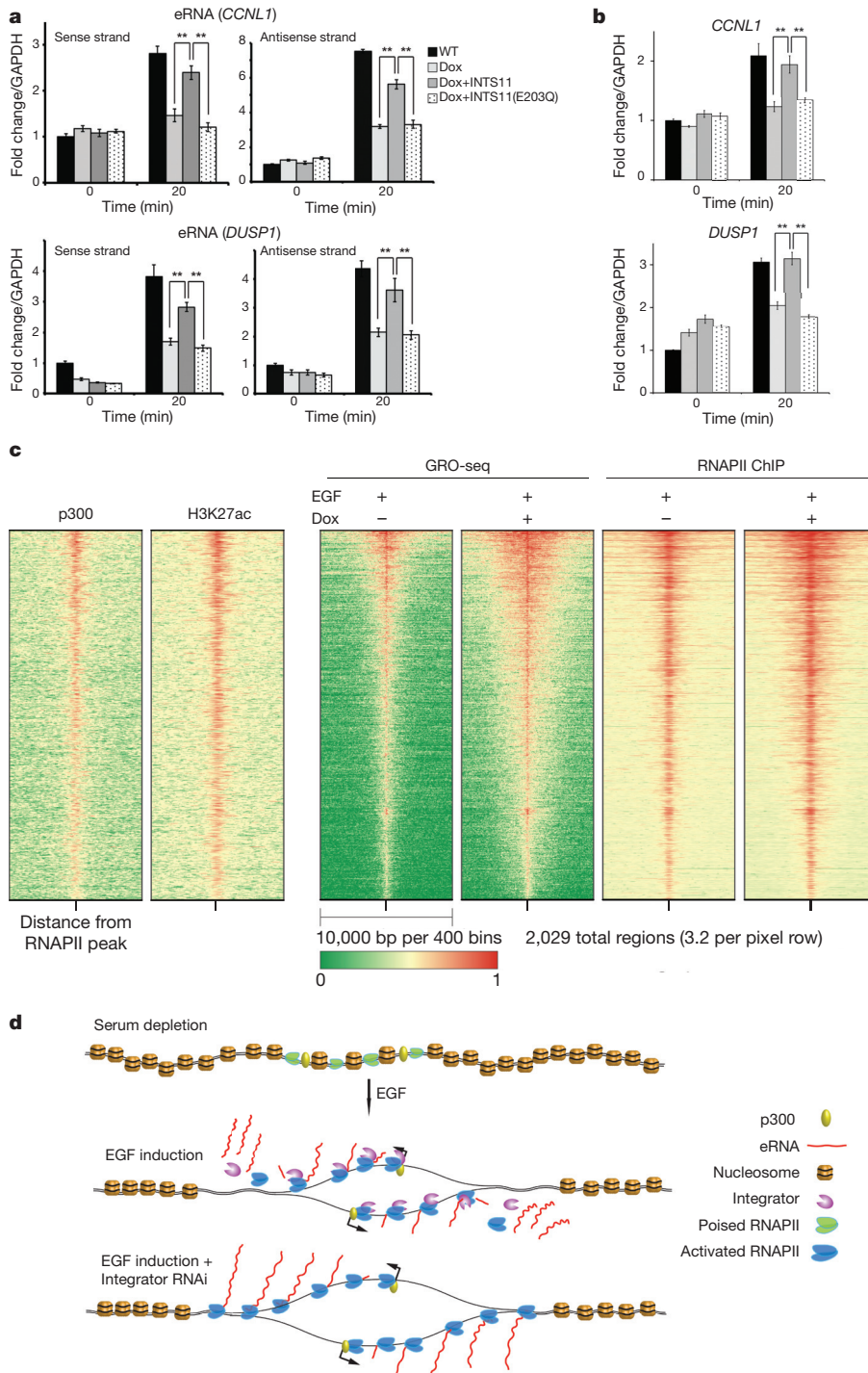


engagement with RNAPII following induction with EGF (Extended Data Fig. 8b–d). We found similar results after analysis of RNAPII interaction with the eRNAs at the *ATF3* super-enhancer (Extended Data Fig. 8e–g). Taken together, these results implicate the Integrator complex in the termination of eRNAs and highlight Integrator's role in the release of eRNA transcripts from transcribing RNAPII.

The catalytic subunit of Integrator is composed of the heterodimer of INTS11 and INTS9 enzymes with close homology to CPSF73 and CPSF100, respectively<sup>20</sup>. We previously showed that a single point mutation (E203Q) in the catalytic domain of INTS11 leads to impaired processing of small nuclear RNAs<sup>1</sup>. To assess the impact of INTS11 enzymatic activity on eRNA biogenesis, we developed wild type and mutant INTS11 (E203Q) that would be refractory to the action of

shRNAs against INTS11, and used these constructs to perform rescue experiments. While ectopic expression of wild-type INTS11 could substantially rescue the EGF-induced eRNA levels after depletion of INTS11, the single-point catalytic mutant was without any effect (Fig. 4a and Extended Data Fig. 9a). Interestingly, we observed a similar rescue of the transcriptional activation of EGF-induced genes by the wild-type INTS11 and not its catalytic mutant (Fig. 4b). These results not only demonstrate the requirement of INTS11 catalytic activity in regulating the induction of eRNAs but also highlight the defect in eRNA processing as a contributing factor in the loss of transcriptional responsiveness.

To determine the scope of Integrator function on active enhancers we analysed the 2,029 transcriptionally active enhancers in HeLa cells.



**Figure 4 | Integrator has a global role in enhancer regulation.** **a**, Ectopic expression of wild-type INTS11, and not its catalytic mutant (E203Q), following Integrator depletion can rescue eRNA induction by EGF. **b**, A similar rescue was observed for wild-type INTS11 on the target protein-coding genes. Real-time PCR analysis was performed on *CCNL1* and *DUSP1* eRNAs and their corresponding mRNAs before and after EGF stimulation. Each eRNA was assayed with two sets of primers. Error bars represent  $\pm$  s.e.m. ( $n = 3$  biological independent experiments),  $**P < 0.01$  by two-sided *t*-test. **c**, The heat map showcases 2,029 enhancer regions identified using RNAPII extragenic loci enriched in H3K27 acetylation (see Methods). Enhancers were centred at the middle of the RNAPII peak and ranked by transcription activity (GRO-seq). The distribution of p300 and H3K27ac are consistent with a group of active enhancers. Upon Integrator depletion, nascent RNA reads and RNAPII profiles spread beyond the normal 3' end of eRNAs. **d**, Model for the role of Integrator at eRNAs. Stimulation of serum-starved cells with EGF triggers recruitment of RNAPII and Integrator to enhancer sites and induces bi-directional transcription of non-polyadenylated eRNAs. Upon EGF stimulation Integrator navigates the enhancers along with RNAPII to promote endonucleolytic cleavage of nascent transcripts, leading to release of the mature eRNAs. Depletion of Integrator elicits a cleavage defect leading to faulty termination, which results in extended eRNA transcripts and accumulation of RNAPII.



We ranked the enhancers based on their transcriptional activity, which mirrored that of RNAPII occupancy (Fig. 4c). Notably, depletion of Integrator resulted in processing defects at all active enhancers, as reflected by the broadening of GRO-seq and RNAPII ChIP-seq reads commensurate with the transcriptional activity of each enhancer site (Fig. 4c). This was in contrast to GRO-seq and RNAPII profiles at transcriptionally active protein-coding genes (Extended Data Fig. 9b). These results demonstrate the generality of Integrator in the processing of eRNAs at enhancers (Fig. 4d).

Recent genome-wide studies have revealed the presence of RNAPII at active enhancers coincident with expression of these regulatory elements as long non-coding RNAs<sup>5,21</sup>. Importantly, such eRNAs have been shown to have critical roles in transcriptional induction by a variety of signal transduction pathways<sup>7,8,11,16,22</sup>. We show that Integrator is the molecular machine that is recruited to enhancers in a signal-dependent manner and is required for the induction of eRNAs. We surmise that the defect in 3'-end processing following Integrator depletion leads to a termination defect reflected in increased levels of primary eRNA transcripts. It is also likely that Integrator affects the stability of the mature transcripts, since its depletion leads to changes in steady-state levels of mature eRNAs.

Similar to other regulatory complexes, Integrator is also recruited to the promoters of protein-coding genes including IEGs<sup>18,23,24</sup>. Interestingly, recent reports described an association between Integrator and transcriptional pause release factors, negative elongation factor (NELF) and SPT4–SPT5 complexes<sup>18,19,23,24</sup>. NELF was also reported to associate with eRNAs in neuronal cells<sup>25</sup>. Indeed, we found that Integrator depletion resulted in a defect in transcriptional initiation as well as pause release, which was reflected in the loss of responsiveness of IEGs to EGF stimulation<sup>18</sup>. However, depletion of NELF subunits did not affect eRNA induction (Extended Data Fig. 7c, d). Moreover, Integrator depletion did not change NELF occupancy at EGF-induced enhancers (Extended Data Fig. 7e). Taken together, our results point to multiple functions for Integrator at protein-coding genes. While Integrator at promoters regulates pause release factors, leading to modulation of productive transcriptional elongation, Integrator at enhancers governs eRNA maturation and enhancer–promoter communication.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 14 March 2014; accepted 7 July 2015.**

**Published online 26 August 2015.**

1. Baillat, D. *et al.* Integrator, a multiprotein mediator of small nuclear RNA processing, associates with the C-terminal repeat of RNA polymerase II. *Cell* **123**, 265–276 (2005).
2. Wang, K. C. *et al.* A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. *Nature* **472**, 120–124 (2011).
3. Ørom, U. A. *et al.* Long noncoding RNAs with enhancer-like function in human cells. *Cell* **143**, 46–58 (2010).

4. De Santa, F. *et al.* A large fraction of extragenic RNA pol II transcription sites overlap enhancers. *PLoS Biol.* **8**, e1000384 (2010).
5. Kim, T. K. *et al.* Widespread transcription at neuronal activity-regulated enhancers. *Nature* **465**, 182–187 (2010).
6. Bhatt, D. M. *et al.* Transcript dynamics of proinflammatory genes revealed by sequence analysis of subcellular RNA fractions. *Cell* **150**, 279–290 (2012).
7. Lam, M. T. *et al.* Rev-Erbs repress macrophage gene expression by inhibiting enhancer-directed transcription. *Nature* **498**, 511–515 (2013).
8. Li, W. *et al.* Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation. *Nature* **498**, 516–520 (2013).
9. Wang, D. *et al.* Reprogramming transcription by distinct classes of enhancers functionally defined by eRNA. *Nature* **474**, 390–394 (2011).
10. Sigova, A. A. *et al.* Divergent transcription of long noncoding RNA/mRNA gene pairs in embryonic stem cells. *Proc. Natl Acad. Sci. USA* **110**, 2876–2881 (2013).
11. Hah, N. *et al.* A rapid, extensive, and transient transcriptional response to estrogen signaling in breast cancer cells. *Cell* **145**, 622–634 (2011).
12. Whyte, W. A. *et al.* Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153**, 307–319 (2013).
13. Lovén, J. *et al.* Selective inhibition of tumor oncogenes by disruption of super-enhancers. *Cell* **153**, 320–334 (2013).
14. Hnisz, D. *et al.* Super-enhancers in the control of cell identity and disease. *Cell* **155**, 934–947 (2013).
15. Sanyal, A., Lajoie, B. R., Jain, G. & Dekker, J. The long-range interaction landscape of gene promoters. *Nature* **489**, 109–113 (2012).
16. Melo, C. A. *et al.* eRNAs are required for p53-dependent enhancer activity and gene transcription. *Mol. Cell* **49**, 524–535 (2013).
17. Mousavi, K. *et al.* eRNAs promote transcription by establishing chromatin accessibility at defined genomic loci. *Mol. Cell* **51**, 606–617 (2013).
18. Gardini, A. *et al.* Integrator regulates transcriptional initiation and pause release following activation. *Mol. Cell* **56**, 128–139 (2014).
19. Yamamoto, J. *et al.* DSIF and NELF interact with Integrator to specify the correct post-transcriptional fate of snRNA genes. *Nat. Commun.* **5**, 4263 (2014).
20. Albrecht, T. R. & Wagner, E. J. snRNA 3' end formation requires heterodimeric association of integrator subunits. *Mol. Cell Biol.* **32**, 1112–1123 (2012).
21. Koch, F. *et al.* Transcription initiation platforms and GTF recruitment at tissue-specific enhancers and promoters. *Nature Struct. Mol. Biol.* **18**, 956–963 (2011).
22. Yang, L. *et al.* lncRNA-dependent mechanisms of androgen-receptor-regulated gene activation programs. *Nature* **500**, 598–602 (2013).
23. Stadelmayer, B. *et al.* Integrator complex regulates NELF-mediated RNA polymerase II pause/release and processivity at coding genes. *Nat. Commun.* **5**, 5531 (2014).
24. Skaar, J. R. *et al.* The Integrator complex controls the termination of transcription at diverse classes of gene targets. *Cell Res.* **25**, 288–305 (2015).
25. Schaukowitz, K. *et al.* Enhancer RNA facilitates NELF release from immediate early genes. *Mol. Cell* **56**, 29–42 (2014).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We would like to thank J. M. Marinis and M. A. Lazar for technical support for GRO-seq experiments. We thank D. Hu in A. Shilatfard's laboratory for performing the SEC ChIP-seq experiments. We thank the Oncogenomics core facility at Sylvester Comprehensive Cancer Center for performing high-throughput sequencing. We also thank Shiekhhattar laboratory members and P.-J. Hamard for support and discussions. This work was supported by funds from University of Miami Miller School of Medicine, Sylvester Comprehensive Cancer Center and grants R01 GM078455 and R01 GM105754 (R.S.) from the National Institute of Health.

**Author Contributions** F.L. and A.G. are co-first authors. R.S., F.L. and A.G. conceived and designed the overall project. F.L., A.G. and A.Z. performed the experiments. R.S., F.L. and A.G. analysed the data and wrote the paper.

**Author Information** High-throughput data are deposited at the Gene Expression Omnibus (GEO) under accession number GSE68401. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to R.S. ([rshekhhattar@med.miami.edu](mailto:rshekhhattar@med.miami.edu)).

## METHODS

No statistical methods were used to predetermine sample size.

**Genome-wide data.** High-throughput sequencing data analysed in this study are originally described in ref. 18 and are deposited at the Gene Expression Omnibus with accession number GSE40632.

H3K27ac, H3K4me1 and p300 data sets from HeLa-S3 cells are available as part of the ENCODE project<sup>26</sup> and can be retrieved under the following accession numbers: GSM733684, GSM798322, GSM93550. Additional experiments are deposited at GEO (GSE68401) and include RNA-seq data (chromatin-bound RNA, polyadenylated and non-polyadenylated fractions of total RNA) as well ChIP-seq experiments (acetylation of H3K27 and occupancy of NELFA). Every genome-wide experiment is performed in two independent biological replicates.

**Genome-wide identification of eRNA loci.** Peak analysis of RNAPII ChIP-seq data after EGF stimulation was performed using HOMER 4.6 (run in 'factor' mode). Next, we used the BEDtools suite to discard any peak overlapping to: (i) all exons from Hg19 UCSC Known Genes (with additional 2 kb surrounding every exon); (ii) RNA Genes (from the Hg18 genome annotation table, plus additional 1 kb); (iii) tRNA Genes (Hg19, plus additional 1 kb). We further selected peaks overlapping ( $\pm 400$  bp) with H3K27ac peaks from the ENCODE ChIP-seq obtained in HeLa-S3 (GEO GSE31477). The analysis resulted in 2,029 regions that were further examined for their transcriptional response to EGF. Briefly, we centred a 6-kb window at the midst of the RNAPII peak and we used HOMER 4.6 to calculate RPKM across the entire eRNA locus using chromRNA-seq data before and after EGF induction. We selected a group of 225 EGF-inducible eRNAs displaying a fold change greater than 2 (ctrl versus EGF) and identified the nearest EGF regulated gene (fold change RPKM  $> 1.6$ ). 91 EGF-induced enhancer RNAs located within 500 kb from the nearest EGF-responsive protein-coding genes were selected for further analysis.

**ChIP-seq data analysis.** ChIP-seq data were obtained using HiSeq 2000 and NextSeq 500. Reads were aligned to the human genome hg19 using bowtie2<sup>27</sup> (end to end alignment, sensitive option). Snapshots of raw ChIP-seq data presented throughout the figures were obtained as follows: BigWiggle files for every ChIP-seq were generated using samtools, bedtools and RseQC<sup>28</sup>, these tracks were then uploaded to the UCSC Genome Browser hg19.

**Clustering, heat maps and average density analysis.** ChIP-seq, GRO-seq and RNA-seq data were subjected to read density analysis; seqMINER 1.3.3<sup>29</sup> was used to extract read densities at all enhancer loci with the following parameters: 5' extension = 4 kb, 3' extension = 4 kb, no read extension, total bin number = 180 bins. Mean density profiles were then generated in R 3.0.1 and normalized to sequencing depth. Heat maps were generated with ChAsE (<http://chase.cs.univie.ac.at/>), using default parameters, a 10 kb window and 400 bins and with *ngsplot*<sup>30</sup>.

**qChIP.** ChIP was performed in HeLa cells as already described<sup>18</sup>. Cells were cross-linked with 1% formaldehyde for 10 min at room temperature, harvested and washed twice with 1× PBS. The pellet was resuspended in ChIP lysis buffer (150 mM NaCl, 1% Triton-X 100, 0.7% SDS, 500  $\mu$ M DTT, 10 mM Tris-HCl, 5 mM EDTA) and chromatin was sheared to an average length of 200–400 bp, using a Bioruptor sonication device (20 min with 30 s intervals). The chromatin lysate was diluted with SDS-free ChIP lysis buffer and aliquoted into single immunoprecipitations of  $2.5 \times 10^6$  cells each. A specific antibody or a total rabbit IgG control was added to the lysate along with Protein A magnetic beads (Invitrogen) and incubated at 4 °C overnight. On day 2, beads were washed twice with each of the following buffers: Mixed Micelle Buffer (150 mM NaCl, 1% Triton-X 100, 0.2% SDS, 20 mM Tris-HCl, 5 mM EDTA, 65% sucrose), Buffer 500 (500 mM NaCl, 1% Triton-X 100, 0.1% Na deoxycholate, 25 mM HEPES, 10 mM Tris-HCl, 1 mM EDTA), LiCl/detergent wash (250 mM LiCl, 0.5% Na deoxycholate, 0.5% NP-40, 10 mM Tris-HCl, 1 mM EDTA) and a final wash was performed with 1× TE. Finally, beads were resuspended in 1× TE containing 1% SDS and incubated at 65 °C for 10 min to elute immunocomplexes. Elution was repeated twice, and the samples were further incubated overnight at 65 °C to reverse cross-linking, along with the untreated input (2.5% of the starting material). After treatment with 0.5 mg ml<sup>-1</sup> proteinase K for 3 h, DNA was purified with Wizard SV gel and PCR Clean-up system (Promega). ChIP eluates and input were assayed by real-time quantitative PCR in a 20  $\mu$ l reaction with the following: 0.4  $\mu$ M of each primer, 10  $\mu$ l of iQ SYBR Green Supermix (BioRAD), and 5  $\mu$ l of template DNA (corresponding to 1/40 of the elution material) using a CFX96 real-time system (BioRAD). Thermal cycling parameters were: 3 min at 95 °C, followed by 40 cycles of 10 s at 95 °C, 20 s at 63 °C followed by 30 s at 72 °C.

**Subcellular fractionation.** Subcellular fractionation was followed as described<sup>6</sup>, with minor changes. The cell lysate was re-suspended in cold lysis buffer with 0.15% NP-40, and the sucrose buffer was used to isolate nuclei. Glycerol buffer (20 mM Tris pH 7.9, 75 mM NaCl, 0.5 mM EDTA, 50% glycerol, 0.85 mM DTT) and nuclei lysis buffer (20 mM HEPES pH 7.6, 7.5 mM MgCl<sub>2</sub>, 0.2 mM EDTA,

0.3 M NaCl, 1 M urea, 1% NP-40, 1 mM DTT) were used to isolate nucleoplasmic fraction and chromatin-bound RNA fraction. Chromatin-bound RNA was isolated with Trizol protocol.

**RNA isolation for high-throughput sequencing.** Total RNA or chromatin-bound RNA was extracted using Trizol reagent (Life Technologies). Genomic DNA and ribosomal RNA was removed with Turbo DNA-free kit and RiboMinus Eukaryote Kit (Life Technologies). The polyA and non-polyA fractions were isolated by running RNA samples three times through the Oligo(dT) Dynabeads (Life technologies) to ensure complete separation. The resulting RNA fractions were subjected to strand-specific library preparation using NEBNext Ultra Directional RNA Library Prep Kit for Illumina (New England Biolabs). Sequencing was performed on Nextseq500 (Illumina).

**ChIP-seq.** ChIP-sequencing was performed as previously described<sup>6</sup>.  $1 \times 10^7$  cells were crosslinked in 1% formaldehyde for 10 min and sonicated with a Bioruptor to obtain chromatin fragments of 200–300 bp. Immunoprecipitation was performed overnight with the specific antibodies and Dynabeads Protein A or Protein G beads (Life Technologies). Beads were washed and chromatin fragments were eluted in TE with 1% SDS at 65 °C. After de-crosslinking overnight, DNA was extracted using Wizard SV extraction columns (Promega) and Illumina sequencing libraries were prepared using NEBNext ChIP-seq library per reagent set (New England Biolabs) and following manufacturer's instructions. Libraries were assayed on a BioAnalyzer (High Sensitivity DNA kit) and sequenced on a Nextseq500 (Illumina).

**Antibodies.** Chromatin immunoprecipitation was performed with polyclonal antibodies against INTS11, INTS9, INTS1 (Bethyl, A301-274A, A300-412A, A300-361A). ChIP-seq of NELFA and H3K27ac were performed with goat polyclonal antibodies (Santa Cruz, sc-23599) and rabbit polyclonal antibodies (Abcam, ab4729), respectively.

Antibodies used for immunoblot analysis were:  $\gamma$ -tubulin (Santa Cruz, mouse monoclonal, sc-17788), CBP80 (Santa Cruz, mouse monoclonal sc-271304), INTS1 (Bethyl, rabbit polyclonal, A300-361A) and a proprietary rabbit polyclonal raised against the C terminus of INTS11. Flag M2-conjugated beads (Sigma, A2220) were used for immunoprecipitation.

**Chromosome conformation capture (3C).** 3C assay was performed as previously described with minor changes<sup>31</sup>. HeLa cells were filtered through a 70  $\mu$ m strainer to obtain single cell preparation.  $1 \times 10^7$  cells were then fixed in 1% formaldehyde for 30 min at room temperature for cross-linking. The reaction was quenched with 0.25 M glycine and cells were collected by centrifugation at 240g for 8 min at 4 °C. Cell pellet was lysed in 5 ml cold lysis buffer (10 mM Tris-HCl, pH 7.5; 10 mM NaCl; 5 mM MgCl<sub>2</sub>; 0.1 mM EGTA) with freshly added protease inhibitors (Roche) on ice for 15 min. Isolated nuclei were collected by centrifugation at 400g for 5 min at 4 °C then re-suspended in 0.5 ml of 1.2× restriction enzyme buffer (NEB) with 0.3% SDS and incubated for 1 h at 37 °C while shaking at 900 r.p.m. Next, samples were incubated for 1 h at 37 °C after addition of 2% (final concentration) Triton X-100. 400 U of restriction enzyme was added to the nuclei and incubated at 37 °C overnight. 10  $\mu$ l of samples were collected before and after the enzyme reaction to evaluate digestion efficiency. The reaction was stopped by addition of 1.6% SDS (final concentration) and incubation at 65 °C for 30 min while shaking at 900 r.p.m. The sample was then diluted 10-fold with 1.15× ligation Buffer (NEB) and 1% Triton X-100 and incubated for 1 h at 37 °C while shaking at 900 r.p.m. 400 U of T4 DNA ligase (NEB) were added to the sample and the reaction was carried at 16 °C for 4 h followed by 30 min at room temperature. For each sample, 300  $\mu$ g of Proteinase K were added for protein digestion and de-crosslinking at 65 °C overnight. On the next day, RNA was removed by adding 300  $\mu$ g of RNase and incubating the sample for 1 h at 37 °C. DNA was purified twice by phenol-chloroform extraction and ethanol precipitation. Purified DNA was then analysed by conventional or quantitative PCR. As control for ligation products, the Bac-clones were digested with 10 U of restriction enzyme overnight and then incubated with 10 U T4 DNA-ligase at 16 °C overnight. The DNA was extracted by phenol-chloroform and precipitated with ethanol. Purified DNA was then analysed by conventional or quantitative PCR. For real-time PCR, the  $\Delta C_t$  method was applied for analysing data, using the Bac-clone  $C_t$  values as control. Primer sequences for PCR are listed in Supplementary Table 2. Bac clone ID: RP-11-294A10, RP-11-1107P14, RP-11-1068G13 (Empire Genomics).

**Pol II RNA immunoprecipitation.** RIP was performed as described<sup>31</sup>. HeLa cells were UV-crosslinked at 254 nm (200 mJ cm<sup>-2</sup>) in 10 ml ice-cold PBS and collected by scraping. Cells were incubated in lysis solution (0.1% SDS, 0.5% NP40, 0.5% sodium deoxycholate, 400 Uml21 RNase Inhibitor (Roche)) and protease inhibitor at 4 °C for 25 min with rotation, followed by DNase treatment (30 U of DNase, 15 min at 37 °C). Protein A Dynabeads (Invitrogen) were incubated with 2  $\mu$ g Pol II antibody (Santa Cruz, N-20) and the cell lysate at 4 °C overnight. The purified

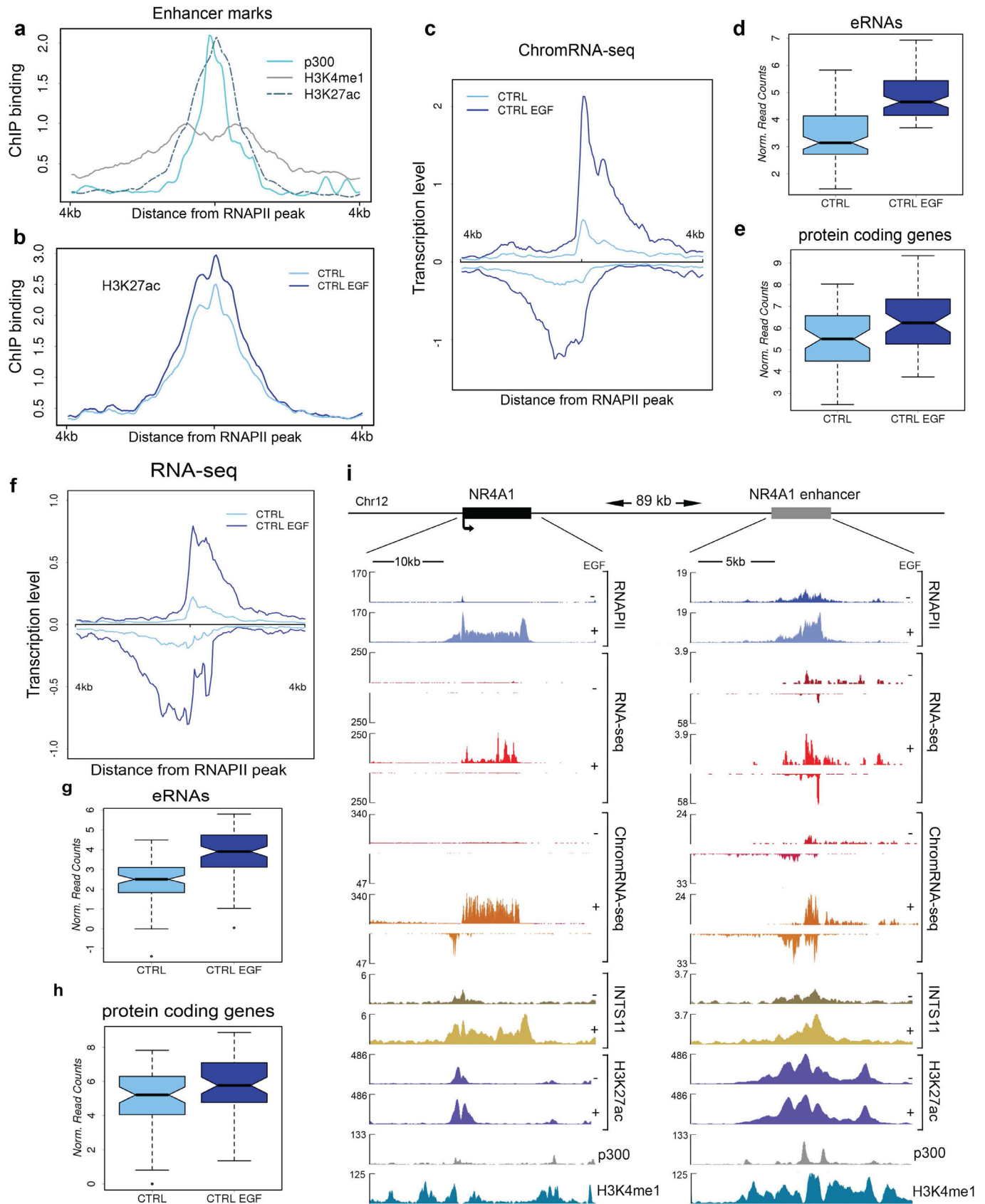
protein–RNA complex was extracted using TRIzol method for RNA extraction and subjected to RT–qPCR with corresponding primers.

**Inducible cell lines.** INTS11 and INTS1 knockdown inducible clones were generated from HeLa cells using the Tet–pLKO–puro vector. For EGF induction, cells were serum starved in 0.5% FBS for 48 h and treated with  $100 \text{ ng ml}^{-1}$  EGF (Invitrogen) for the indicated time course. All cell lines in this study are mycoplasma negative.

**Transfections.** Cells were treated with doxycycline for 48 h. 24 h before EGF induction, INTS11 and INTS11 (E203Q) mutant protein expression plasmids were transfected using Lipofectamine 2000 (Life Technologies, Inc.) according to the manufacturer's instruction. Cells were harvested 0 and 20 min after EGF induction.

All the PCR primer sequences are listed in the supplementary Table 2.

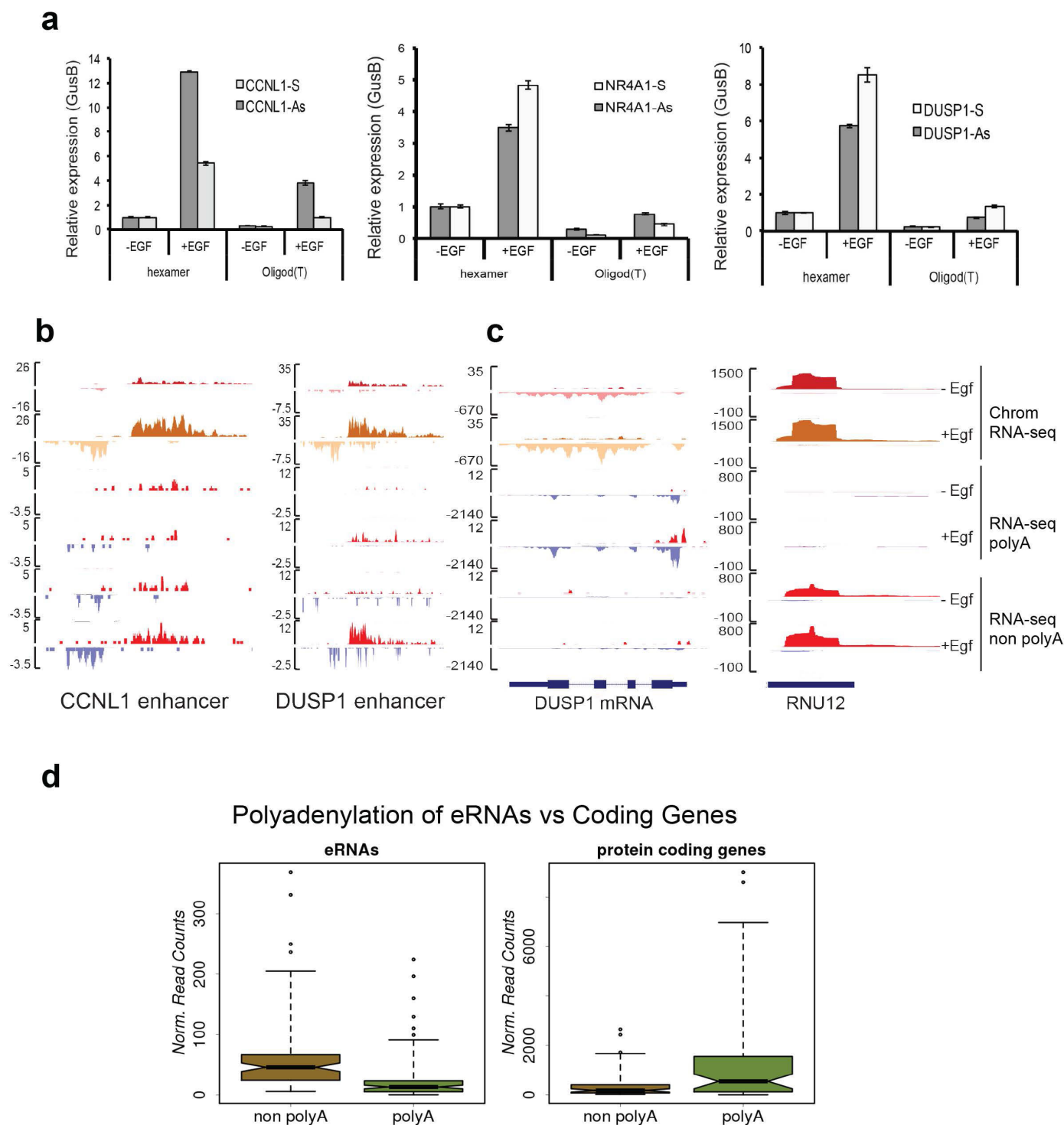
26. Gerstein, M. B. *et al.* Architecture of the human regulatory network derived from ENCODE data. *Nature* **489**, 91–100 (2012).
27. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
28. Wang, L. G., Wang, S. Q. & Li, W. RSeQC: quality control of RNA-seq experiments. *Bioinformatics* **28**, 2184–2185 (2012).
29. Ye, T. *et al.* seqMINER: an integrated ChIP-seq data interpretation platform. *Nucleic Acids Res.* **39**, e35 (2011).
30. Shen, L., Shao, N., Liu, X. & Nestler, E. ngs.plot: Quick mining and visualization of next-generation sequencing data by integrating genomic databases. *BMC Genomics* **15**, 284 (2014).
31. Lai, F. *et al.* Activating RNAs associate with Mediator to enhance chromatin architecture and transcription. *Nature* **494**, 497–501 (2013).





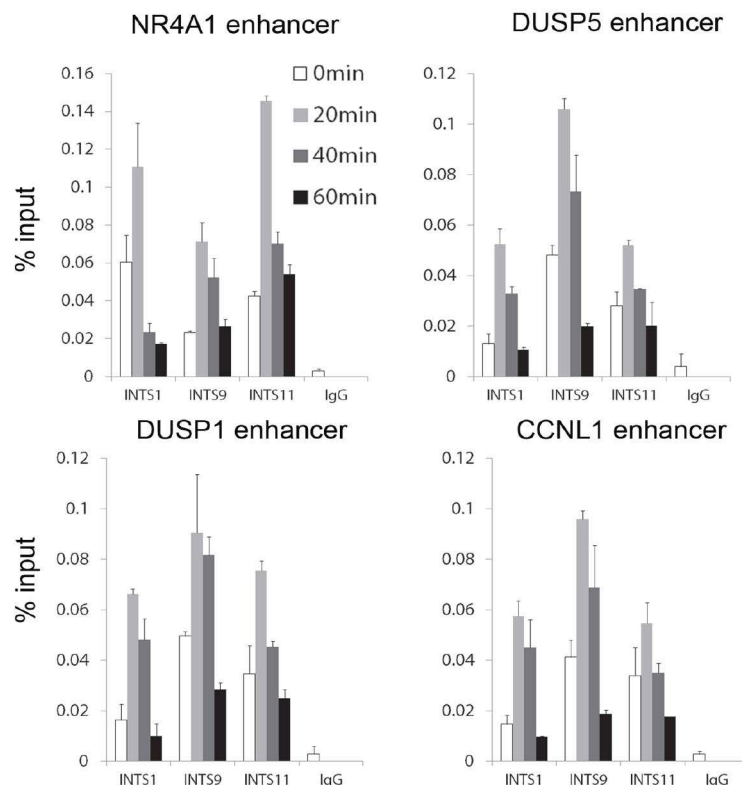
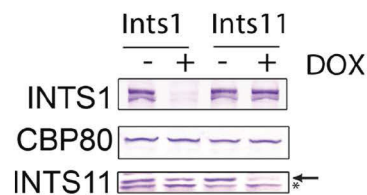
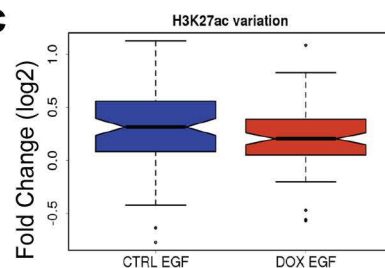
**Extended Data Figure 1 | Identification eRNAs responsive to EGF.** **a**, We identified 91 EGF-responsive enhancer regions in HeLa cells. We annotated extragenic RNAPII sites (see Methods) and used the middle of the RNAPII peak as an anchor to display average profiles of p300, H3K27ac and H3K4me1 (data from the ENCODE project). The profiles represent the mean read density of ChIP-seq data. The 91 loci display a typical enhancer signature, with enrichment of p300 and H3K27ac around the TSS and a broader decoration by H3K4me1. **b**, Profiles of H3K27ac were obtained from ChIP-seq analysis of HeLa cells before and after 20 min of EGF induction. Mean read density was normalized to sequencing depth. **c**, EGF stimulates bi-directional transcription from 91 enhancer regions. We displayed the mean read density obtained from strand-specific sequencing of the chromatin-bound RNA fraction (ChromRNA-seq). **d**, **e**, Normalized read density (RPKM) was calculated from RNA-seq data for 91 eRNAs (**d**) and 57 neighbouring protein-coding genes

(**e**) that responded to EGF stimulation ( $FC > 1.6$ ) and mapped within 500 kb from an EGF-responsive eRNA. **f**, Average profiles of ChromRNA-seq data at 91 enhancer loci (mean density of reads, normalized to total read number). **g**, **h**, Box plot of 91 eRNAs before and after treatment with EGF shows the average increase of transcription 20 min after stimulation ( $P < 0.001$ ), matched by an increase in the neighbouring protein-coding genes ( $P < 0.02$ ). **i**, *NR4A1* is activated by EGF in HeLa cells: RNAPII and INTS11 are recruited to the *NR4A1* locus after 20 min of stimulation, with concomitant accumulation of reads from RNA-seq and ChromRNA-seq. A neighbouring eRNA locus also exhibits increased transcription along with RNAPII and INTS11 recruitment. Sequencing tracks are visualized in BigWig format and aligned to the hg19 assembly of the UCSC Genome Browser. Whiskers on the box plots indicate the variability in the datasets.



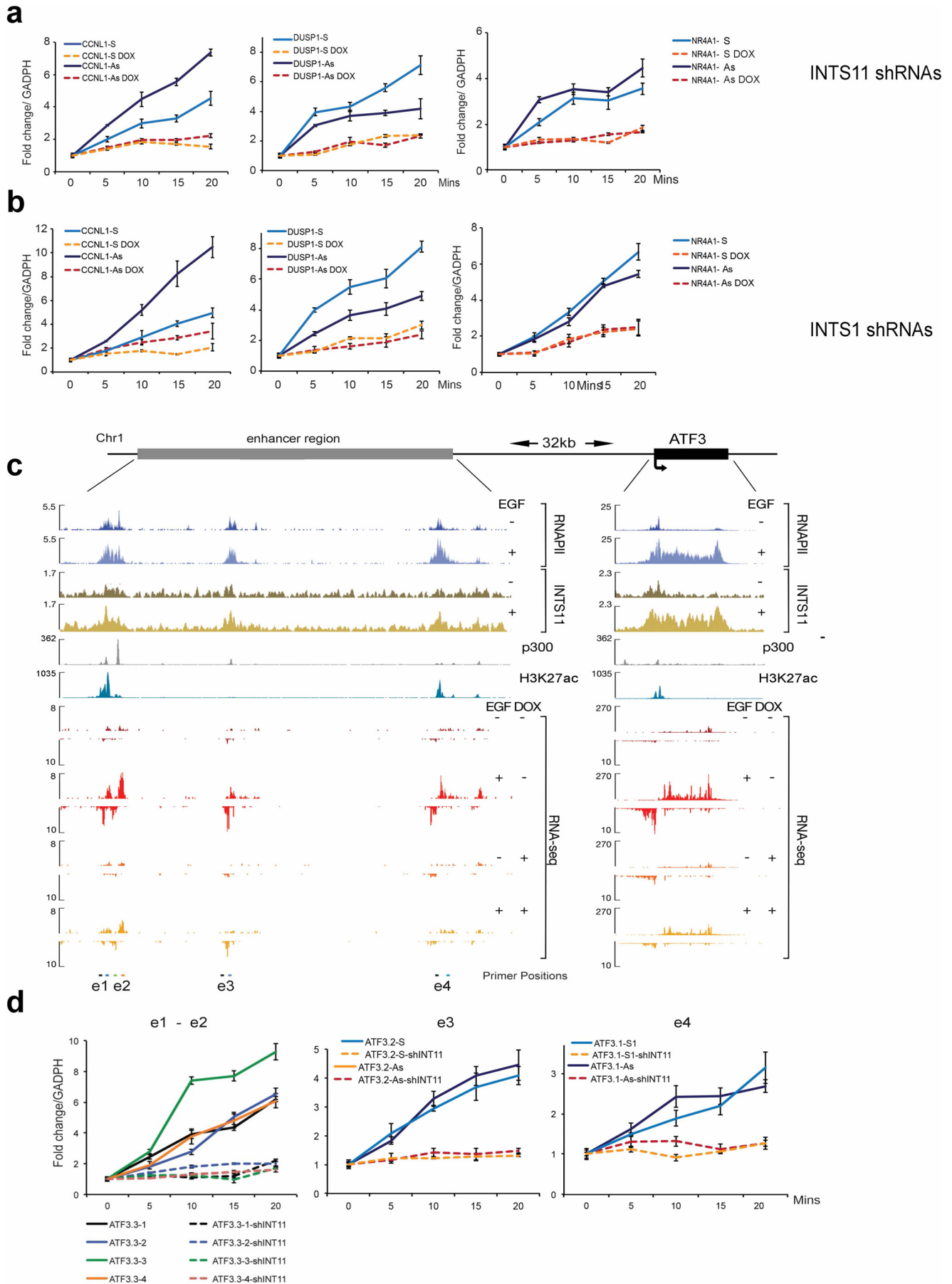
**Extended Data Figure 2 | EGF-induced eRNAs are predominantly non-polyadenylated.** **a**, We examined transcription at three enhancers adjacent to EGF-responsive genes *CCNL1*, *NR4A1* and *DUSP1*. Total RNA samples were collected before and after EGF induction. Reverse transcription was performed with random hexamer primer or oligo d(T) primer. Each eRNA strand was analysed by real-time PCR with specific primers. Error bars represent  $\pm$  standard error of the mean (s.e.m.,  $n = 3$  biological independent experiments).  $P < 0.01$  by two-sided  $t$ -test. **b, c**, RNA-seq was performed on the polyadenylated and non-polyadenylated fraction of total RNA. RNA-seq tracks were visualized in BigWig format and aligned to the hg19 assembly

of the UCSC Genome Browser. *CCNL1* and *DUSP1* enhancers were displayed (**b**) along with a polyadenylated control (*DUSP1* protein-coding locus) and a non-polyadenylated transcript (snRNA U12) (**c**). All EGF-induced eRNAs and protein-coding genes (RefSeq hg19) were examined for their average RPKM throughout the entire locus. **d**, We compared polyadenylation levels of 225 eRNAs and 150 protein-coding genes (2 fold induction upon EGF, RPKM calculated from ChromRNA-seq data previously described). The box plot shows predominance of non-polyadenylated transcripts mapping to eRNA loci, as opposed to transcripts coding for RefSeq genes. Whiskers on the box plot indicate the variability in the datasets.

**a****b****c**

**Extended Data Figure 3 | The Integrator complex is recruited to enhancers upon EGF stimulation.** **a**, qChIP analysis of Integrator occupancy using INTS11, INTS1 and INST9 antibodies at four eRNA loci. Data were collected during a time course of EGF induction in HeLa cells (0, 20, 40 and 60 min). Error bars represent  $\pm$  standard error of the mean (s.e.m.,  $n = 3$  biological independent experiments).  $P < 0.01$  by two-sided  $t$ -test. **b**, Depletion of INST1 and INST11 protein levels in tet-inducible HeLa clones. The arrow indicates the

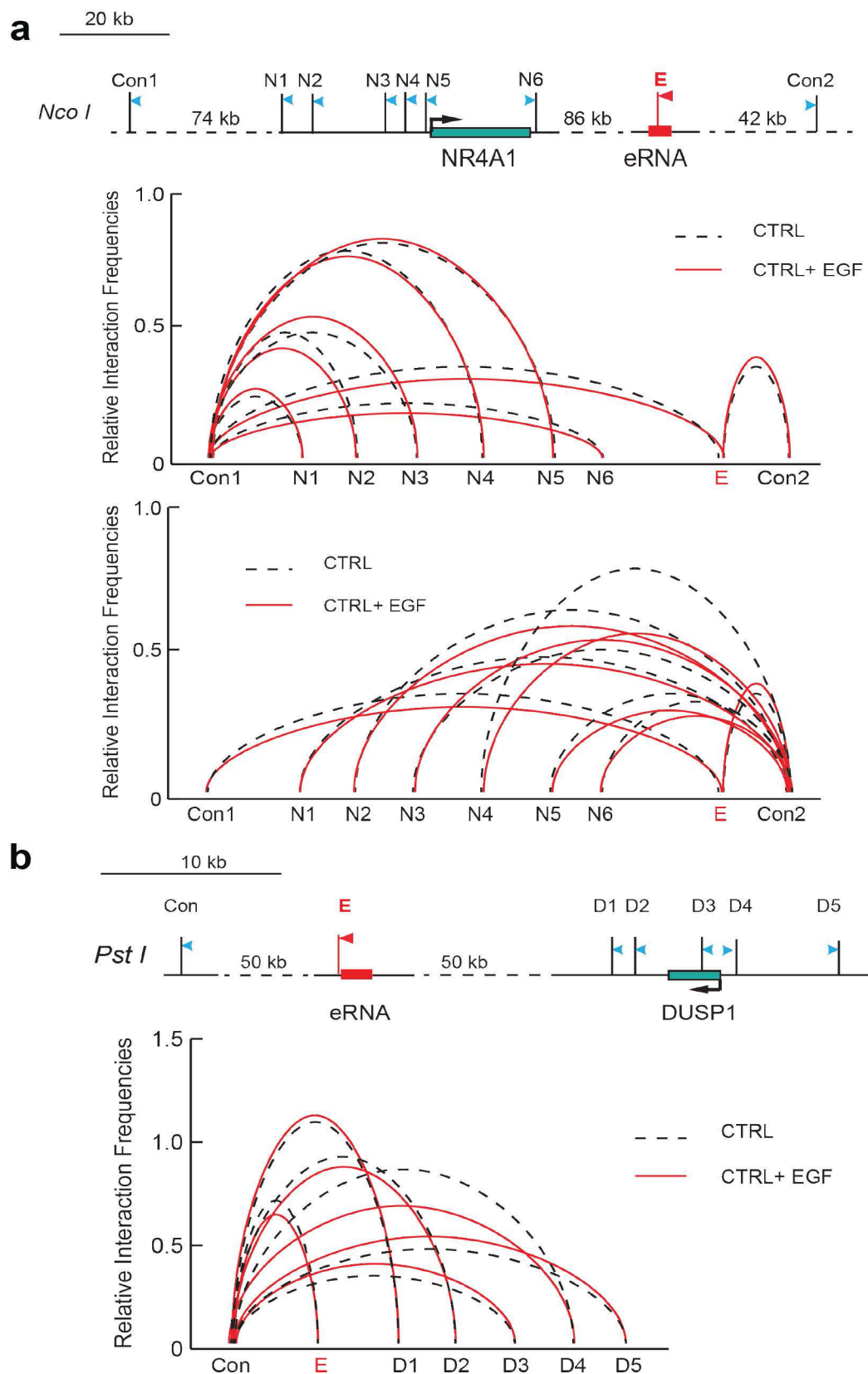
INTS11-specific signal; the asterisk shows a non-specific band. **c**, Fold change of H3K27 acetylation (0 min/20 min EGF) before (ctrl) and after (dox) depletion of INTS11. Data were calculated from read density of ChIP-seq experiments across EGF-induced enhancers. Depletion of Integrator significantly affects EGF-dependent increase in H3K27ac ( $P < 0.05$ ). Whiskers on the box plot indicate the variability in the datasets.





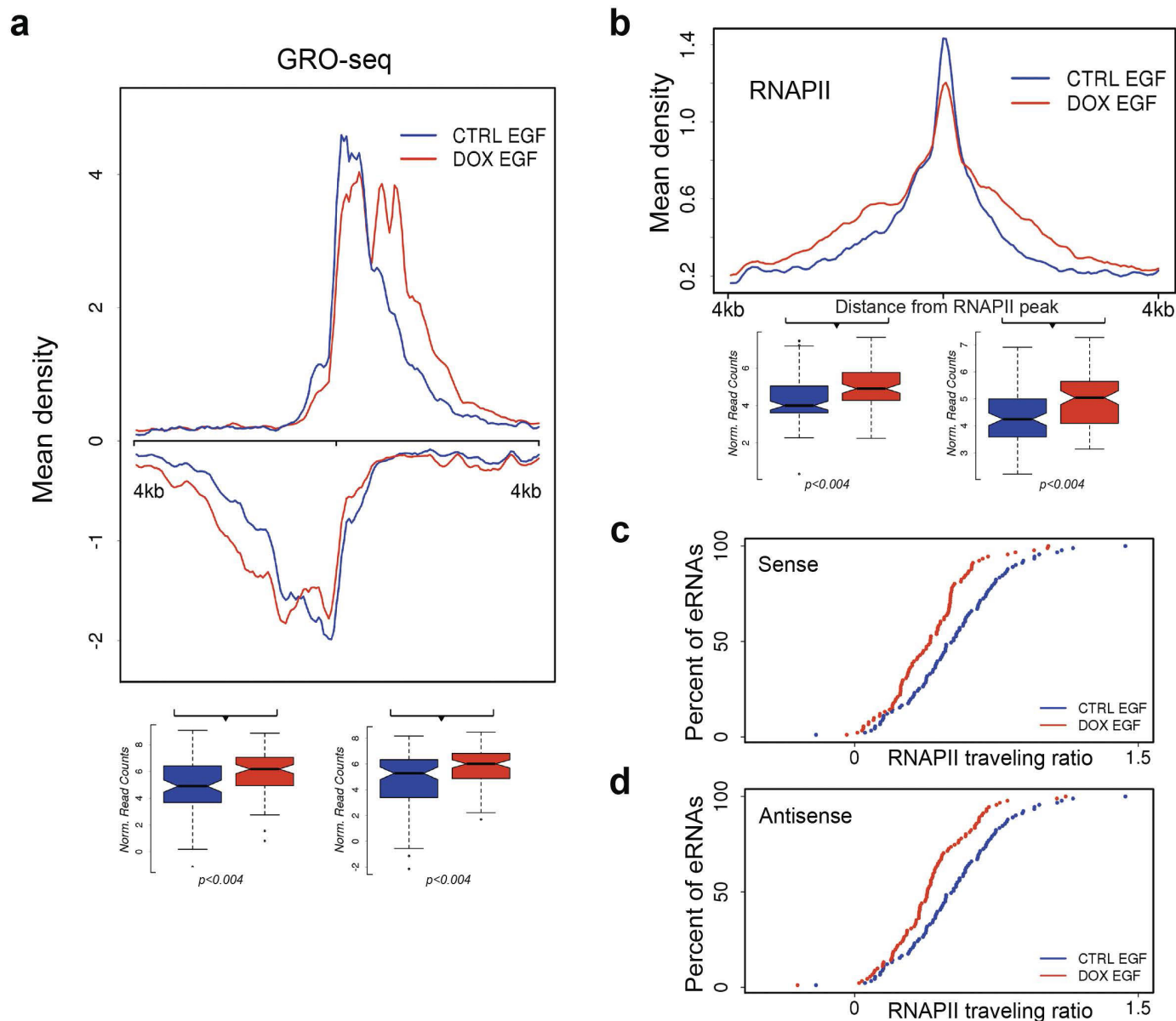
**Extended Data Figure 4 | Depletion of Integrator impairs activation of eRNAs by EGF.** **a, b,** Activation of eRNAs near *DUSP1*, *CCNL1* and *NR4A1* genes were assayed by qRT-PCR in three independent experiments, using INTS11 (**a**) or INTS1 (**b**) inducible shRNA clones. Transcription was followed throughout a 20-min time-course experiment. Each eRNA was amplified with two different sets of specific primers to analyse both strands; dashed lines indicate treatment with doxycycline (dox) to induce shRNAs. Data at every time point are reported as fold change (EGF/non-induced). Error bars represent  $\pm$  s.e.m. ( $n = 3$  biological independent experiments),  $P < 0.01$  by two-sided  $t$ -test. **c,** Schematic representation of *ATF3* and its super-enhancer

region located 30 kb upstream (top). Snapshots of ChIP-seq and RNA-seq tracks show EGF-dependent recruitment of RNAPII and INTS11 at the *ATF3* locus and at several upstream enhancers. Depletion of INTS11 nearly abolished transcription of eRNAs and *ATF3* mRNA. **d,** Real-time RT-PCR analysis of the *ATF3* super-enhancer region upon depletion of INTS11. qPCR analysis was performed before and 5, 10, 15, 20 min after EGF treatment with strand-specific primer sets (indicated below the RNA-seq tracks in **c**). Error bars represent  $\pm$  s.e.m. ( $n = 3$  biological independent experiments),  $P < 0.01$  by two-sided  $t$ -test.



**Extended Data Figure 5 | Chromatin conformation capture at control loci.**  
**a**, 3C analysis of *NR4A1* promoter and control sites. The Con1 site lies 74 kb upstream of the *NR4A1* protein-coding gene and the Con2 site is located 42 kb downstream of the enhancer site. There are no looping events between

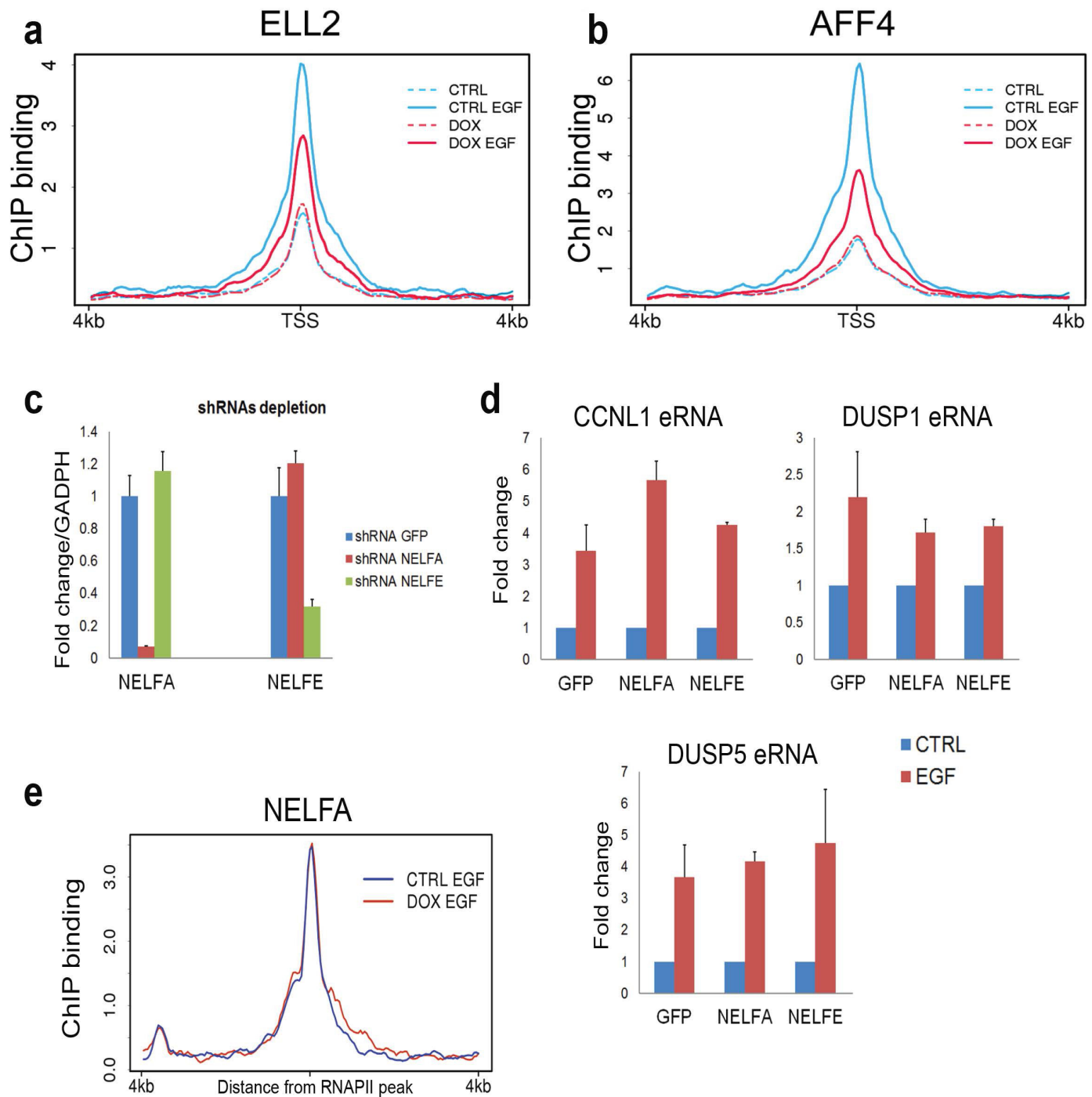
either control sites with the *NR4A1* promoter region after EGF induction.  
**b**, Similarly, no looping events were detected between the promoter of *DUSP1* and a downstream control site (Con). All data were averaged from three independent experiments,  $P < 0.01$  by two-sided  $t$ -test.



#### Extended Data Figure 6 | Integrator has a role in eRNA termination.

**a**, Mean density profiles of GRO-seq data at 91 EGF-induced enhancers. Data are presented as strand-specific mean read density, centred at the middle of the RNAPII peak and normalized to sequencing depth. The underlying box plots were used to quantify the enrichment of GRO-seq reads at the 3' end of both eRNA transcripts (2 kb window, centred 1 kb downstream of the RNAPII peak). **b**, RNAPII profiling at 91 enhancers after INTS11 depletion shows accumulation of ChIP-seq reads towards the 3' end. Data are presented as mean read density, centred at the middle of the RNAPII peak and normalized

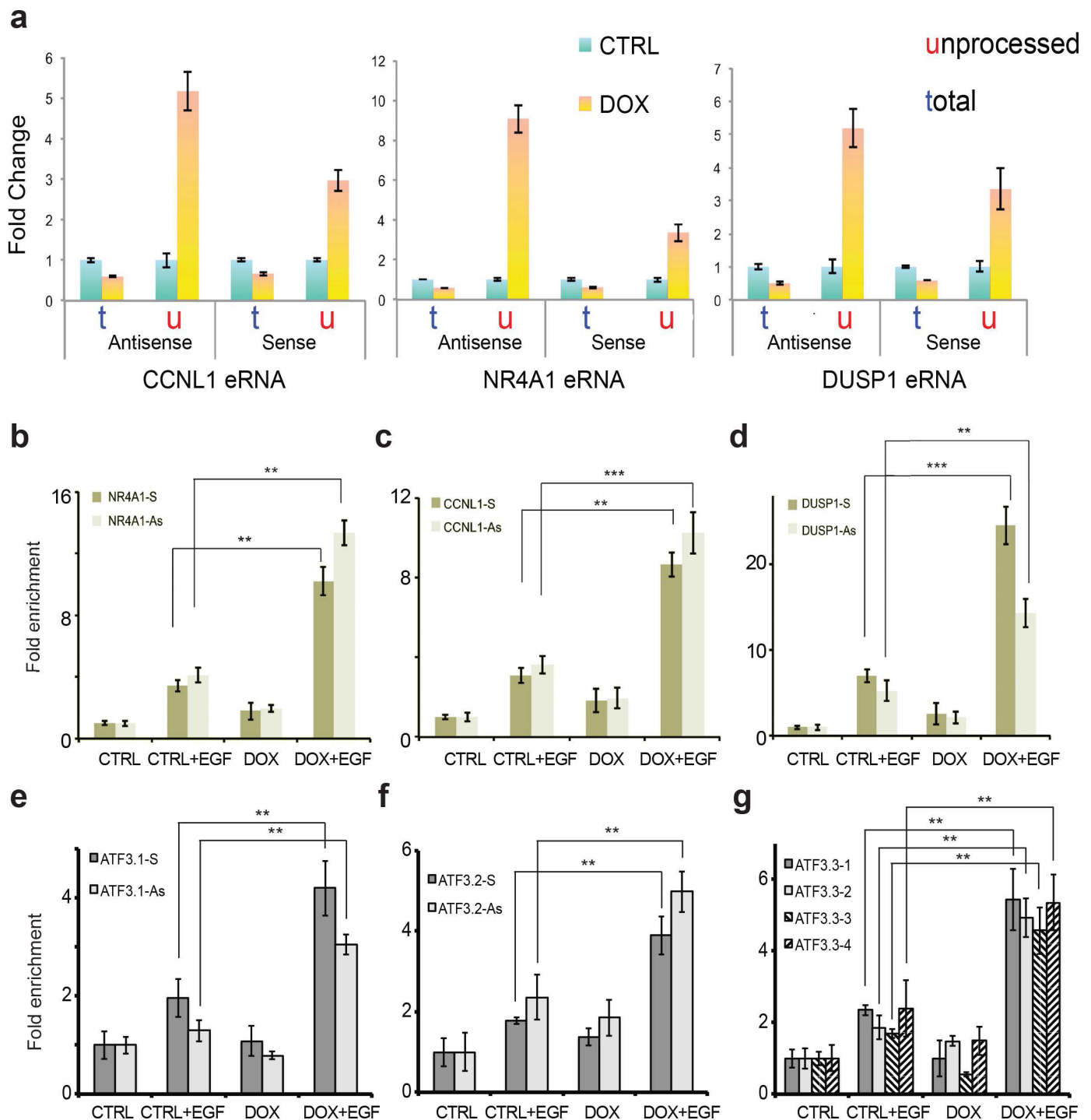
to sequencing depth. Box plots represent the enrichment of RNAPII reads of both eRNA transcripts (2 kb window, centred 1 kb downstream of the RNAPII peak). RNAPII significantly accumulated ( $P < 0.004$ ) after depletion of INTS11. Whiskers on the box plots indicate the variability in the datasets. **c, d**, RNAPII travelling ratio at enhancers was measured as the ratio between RNAPII density close to the transcription start site (the surrounding 300 bp) and 3 kb downstream. Given the bi-directional nature of transcription at enhancers, travelling ratio was calculated for both sense (**c**) and antisense (**d**) transcripts.



**Extended Data Figure 7 | Analysis of super elongation complex at enhancers.** **a, b**, Metagene analysis on 91 eRNA loci shows the effect of EGF stimulation and INTS11 depletion on the recruitment of the ELL2 (**a**) and AFF4 (**b**) subunits of the super elongation complex (SEC). SEC was recruited to enhancers upon EGF stimulation. Depletion of Integrator decreases AFF4 and ELL2 recruitment. Data were visualized as mean read density, normalized to sequencing depth, across 8 kb surrounding the centre of enhancers. **c**, To investigate the role of the negative elongation factor (NELF) in induction of eRNAs, we infected HeLa cells with lentiviral shRNAs against NELFA, NELFE and a control GFP. Quantitative RT-PCR analysis shows the extent of NELF

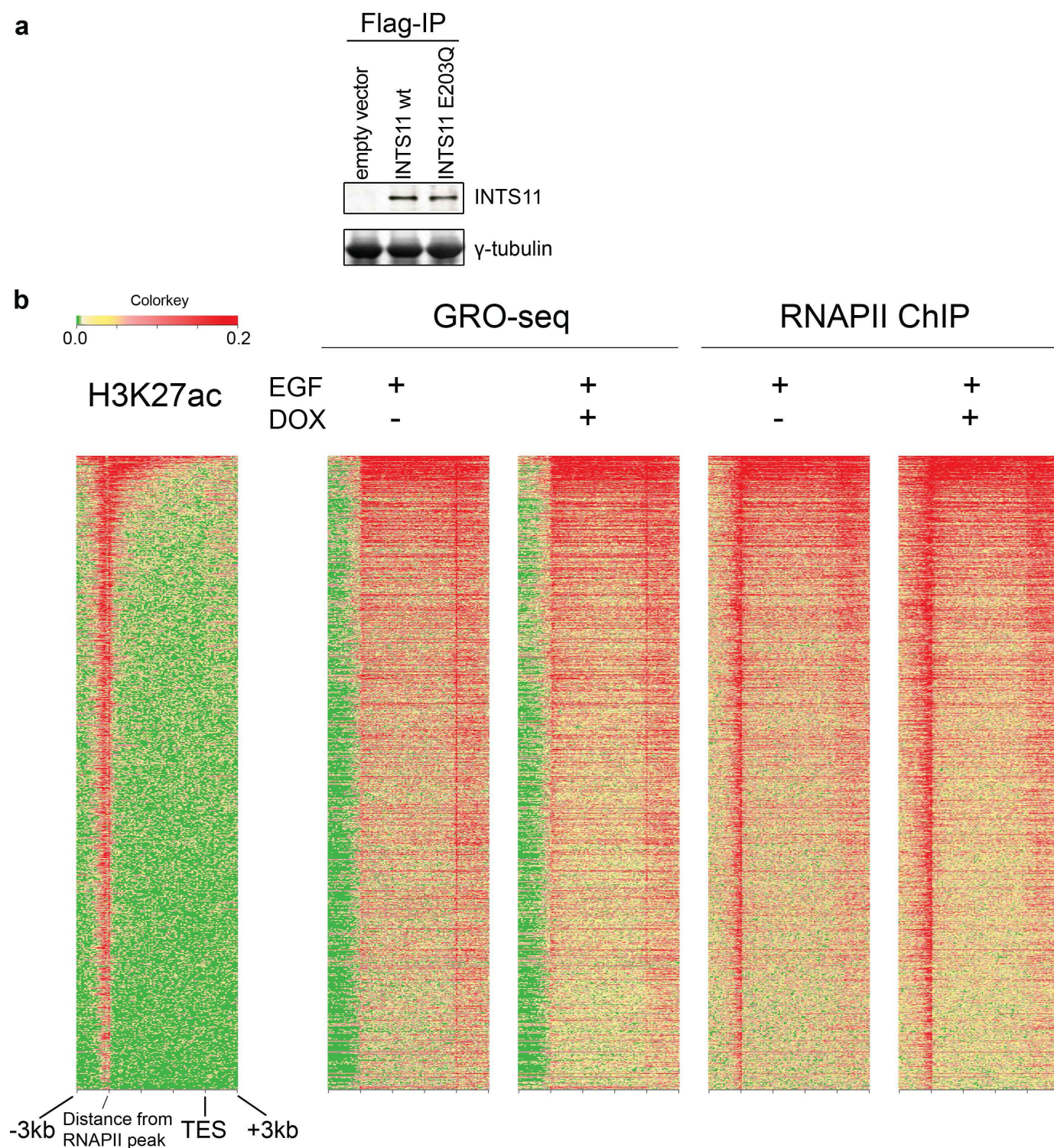
depletion 72 h after infection. Error bars represent  $\pm$  s.e.m. ( $n = 3$  biological independent experiments),  $P < 0.01$  by two-sided  $t$ -test. **d**, Depletion of two different NELF subunits does not significantly impact activation of EGF-responsive eRNAs. Data represent fold change of induction (EGF/not induced) after 20 min of stimulation and were normalized against *GUSB* expression. Error bars represent  $\pm$  s.e.m. ( $n = 3$  biological independent experiments),  $P < 0.01$  by two-sided  $t$ -test. **e**, ChIP-seq analysis of NELFA before and after depletion of INTS11. Metagene analysis shows mean read density (normalized to sequencing depth) across 91 eRNAs. NELF occupancy at enhancers was not affected by depletion of Integrator.





**Extended Data Figure 8 | Integrator depletion causes accumulation of unprocessed eRNAs and prevents release of RNAPII.** **a**, Termination of eRNAs was examined with quantitative RT-PCR. Primer pairs were designed to amplify a portion of the enhancer transcript detected in normal condition (t, total) or a longer template further extending into the 3' of the enhancer region (u, unprocessed). qPCR analysis was performed before (ctrl) and after (dox) depletion of INTS11 at three eRNAs (sense and antisense strand), after stimulation with EGF. In the absence of INTS11, we observed accumulation of unprocessed eRNA, suggestive of a termination defect. Error bars represent  $\pm$  s.e.m. ( $n = 3$  biological independent experiments),  $P < 0.01$  by two-sided  $t$ -test. Release of eRNA transcripts from RNA polymerase was investigated by means of RNAPII immunoprecipitation following UV

cross-link (UV-RIP). **b-d**, After RNAPII immunoprecipitation, eRNAs near *DUSP1*, *CCNL1* and *NR4A1* genes were assayed by qRT-PCR and showed increased association with RNAPII in the absence of Integrator. Each eRNA was detected by two different sets of specific primers (sense and antisense). Error bars represent  $\pm$  s.e.m. ( $n = 3$  biological independent experiments).  $*P < 0.01$ ,  $**P < 0.01$ ,  $***P < 0.001$  by two-sided  $t$ -test. **e-g**, RNAPII UV-RIP analysis was also performed on several eRNAs from the ATF3 super-enhancer. qRT-PCR on the RNA recovered after immunoprecipitation shows increased association between RNAPII and eRNAs in the absence of Integrator. Each eRNA was detected by two different sets of specific primers (sense and antisense). Error bars represent  $\pm$  s.e.m. ( $n = 3$  three independent experiments).  $**P < 0.01$  by two-sided  $t$ -test.



**Extended Data Figure 9 | Distribution of RNAPII and nascent RNAs across protein-coding genes.** **a**, Expression level of exogenous INTS11 wild type (WT) and its catalytic mutant (E203Q). Nuclear extracts were subjected to Flag immunoprecipitation and probed with a polyclonal antibody raised against the C terminus of INTS11. **b**, Heat map of nascent RNA (GRO-seq) and

RNAPII ChIP-seq across the 2,000 most active genes in HeLa cells. Gene loci were analysed for their entire gene body, with 3 additional kilobases on both ends. H3K27ac data from ENCODE is shown on the left; genes are ranked according to the intensity of RNAPII signal. Depletion of Integrator does not appear to affect termination at protein-coding genes.



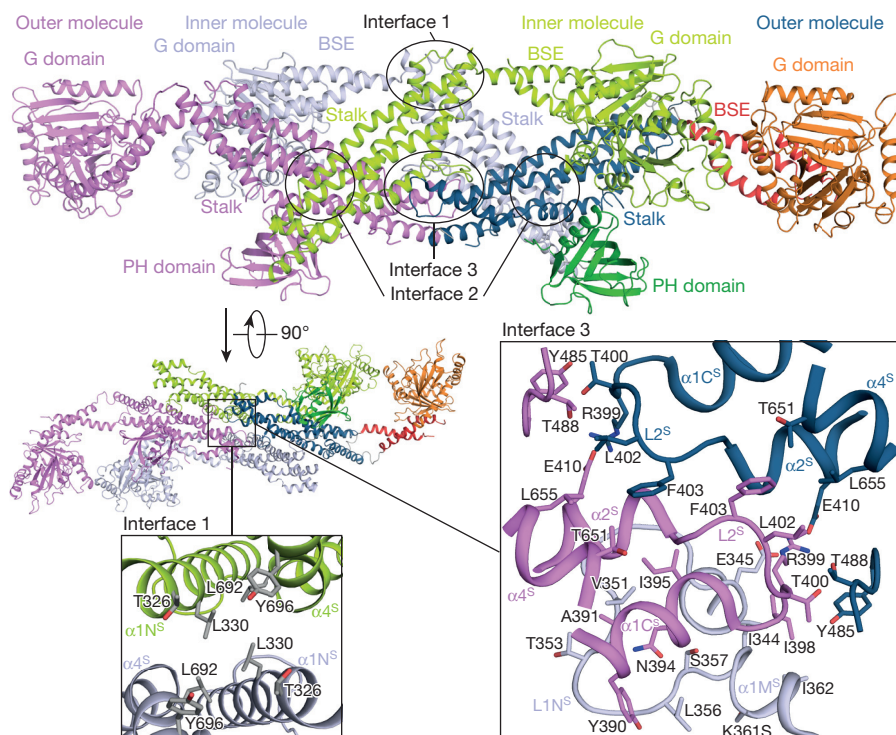
# Crystal structure of the dynamin tetramer

Thomas F. Reubold<sup>1\*</sup>, Katja Faelber<sup>2\*</sup>, Nuria Plattner<sup>3</sup>, York Posor<sup>4†</sup>, Katharina Ketel<sup>4</sup>, Ute Curth<sup>1,5</sup>, Jeanette Schlegel<sup>2</sup>, Roopsee Anand<sup>1</sup>, Dietmar J. Manstein<sup>1,5</sup>, Frank Noé<sup>3</sup>, Volker Haucke<sup>4,6</sup>, Oliver Daumke<sup>2,6</sup> & Susanne Eschenburg<sup>1</sup>

The mechanochemical protein dynamin is the prototype of the dynamin superfamily of large GTPases, which shape and remodel membranes in diverse cellular processes<sup>1</sup>. Dynamin forms predominantly tetramers in the cytosol, which oligomerize at the neck of clathrin-coated vesicles to mediate constriction and subsequent scission of the membrane<sup>1</sup>. Previous studies have described the architecture of dynamin dimers<sup>2,3</sup>, but the molecular determinants for dynamin assembly and its regulation have remained unclear. Here we present the crystal structure of the human dynamin tetramer in the nucleotide-free state. Combining structural data with mutational studies, oligomerization measurements and Markov state models of molecular dynamics simulations, we suggest a mechanism by which oligomerization of dynamin is linked to the release of intramolecular autoinhibitory interactions. We elucidate how mutations that interfere with tetramer formation and autoinhibition can lead to the congenital muscle disorders Charcot–Marie–Tooth neuropathy<sup>4</sup> and centronuclear myopathy<sup>5</sup>, respectively. Notably, the bent shape of the tetramer explains how dynamin assembles into a right-handed helical oligomer of defined

diameter, which has direct implications for its function in membrane constriction.

The three highly conserved vertebrate isoforms of dynamin contain five distinct domains (Extended Data Fig. 1a): an N-terminal GTPase (G) domain mediating nucleotide binding and hydrolysis, a bundle signalling element (BSE), a stalk, a pleckstrin homology (PH) domain involved in lipid binding, and a proline-rich domain (PRD) mediating interactions with scaffolding proteins containing BAR- and SH3-domains<sup>6</sup>. To exert its function in clathrin-mediated endocytosis (CME), dynamin assembles via the stalks into a helical array surrounding the necks of invaginating clathrin-coated pits<sup>7,8</sup>. Dimerization of GTP-bound G domains from neighbouring helical rungs induces GTP hydrolysis<sup>9</sup>. The ensuing conformational changes are thought to be transmitted from the G domain via the BSE to the stalk, resulting in a sliding motion of adjacent helix rungs, concomitant helix constriction<sup>10</sup>, and eventually membrane scission. The inherent tendency to form large assemblies at high protein concentrations has hampered crystallization of dynamin in the past. Previously, the use of non-oligomerizing mutants led to the determination of dynamin 1 crystal



**Figure 1 | Structure of the dynamin 3 tetramer.** The four molecules in the tetramer are coloured separately; in the molecule on the right each domain is individually coloured. The tetramer consists of two dimers, each formed via the

central interface 2. The two dimers are connected via interfaces 1 (left box) and 3 (right box) to build the tetramer. One inner molecule is omitted from the detailed view for clarity.

<sup>1</sup>Institut für Biophysikalische Chemie, Medizinische Hochschule Hannover, Carl-Neuberg-Str. 1, 30625 Hannover, Germany. <sup>2</sup>Max-Delbrück-Centrum für Molekulare Medizin, Kristallographie, Robert-Rössle-Straße 10, 13125 Berlin, Germany. <sup>3</sup>Institut für Mathematik, Freie Universität Berlin, Arnimallee 6, 14195 Berlin, Germany. <sup>4</sup>Leibniz-Institut für Molekulare Pharmakologie, Robert-Rössle-Straße 10, 13125 Berlin, Germany. <sup>5</sup>Forschungseinrichtung Strukturanalyse, Medizinische Hochschule Hannover, Carl-Neuberg-Str. 1, 30625 Hannover, Germany. <sup>6</sup>Institut für Chemie und Biochemie, Freie Universität Berlin, Takustraße 6, 14195 Berlin, Germany. <sup>†</sup>Present address: Cancer Institute, University College London, 72 Huntley Street, London WC1E 6DD, UK.

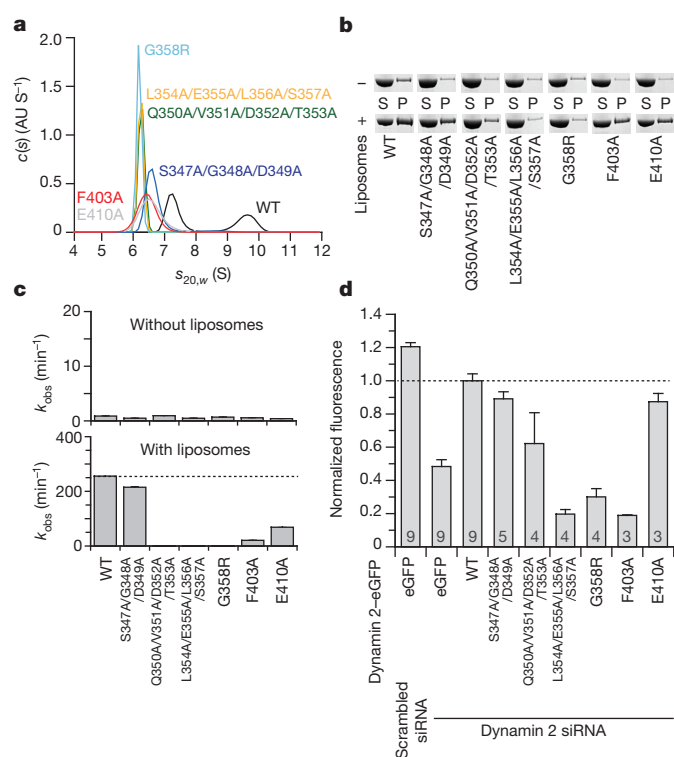
\*These authors contributed equally to this work.

structures<sup>2,3</sup>; however, the postulated higher-order assembly interface was not resolved in these structures, leaving the oligomerization mechanism unaddressed. We reasoned that an alternative assembly-affecting mutation, such as K361S in dynamin 3 (ref. 11), might disturb the oligomerization interface to a lesser extent than the previously used mutants. We obtained crystals of nucleotide-free dynamin 3 (K361S) lacking the PRD (dynamin 3 (K361S/ $\Delta$ PRD)) that diffracted to 3.7 Å (Methods, Extended Data Fig. 1, Extended Data Table 1). The asymmetric unit of the crystal lattice contained a dynamin tetramer that did not form the filamentous superstructures seen for dynamin 1 (refs 2, 3).

The dynamin tetramer consists of two dimers, each of which assembles via the previously described interface 2 (refs 2, 3, 12–14) (Fig. 1, Extended Data Fig. 2a, b). Different dimerization and assembly models were derived from electron microscopy (EM) reconstructions and cross-linking experiments<sup>9</sup>. These models, however, are not compatible with the architecture of the dynamin tetramer (Extended Data Fig. 2c). To provide further evidence for dimerization via interface 2, we introduced the triple mutation I481D/H677D/L678S in dynamin 3 ( $\Delta$ PRD) and the corresponding mutation (I481D/H687D/L688S) in dynamin 1 ( $\Delta$ PRD). These mutants were monomeric in analytical ultracentrifugation experiments (Extended Data Fig. 2d). Thus, dimerization via interface 2 is indeed a general feature of dynamin and dynamin-like proteins. This conclusion receives additional support from recent cross-linking data<sup>15</sup>.

Dynamin 3 dimers further assemble into tetramers via interface 1 and interface 3 (Fig. 1 and Extended Data Fig. 3a). Interface 1 at the top of the stalk features four hydrophobic residues that are highly conserved in the dynamin superfamily (Fig. 1 and Supplementary Fig. 1). The main contributors for interface 3 are loop L1N<sup>S</sup> (superscript S denotes belonging to the stalk) of the 'inner' and loop L2<sup>S</sup> of the 'outer' stalks (Fig. 1), which mediate an intricate interaction network involving all four stalks (Extended Data Fig. 3a). Accordingly, these loop regions are well defined in the inter-dimer interface (Extended Data Fig. 3b), but not at the outer, non-assembled sides of the tetramer. Previous studies have shown that mutation of R399 in loop L2<sup>S</sup> completely destroys higher-order assembly and dynamin function<sup>2,3,11</sup>. In our structure, R399 of an outer molecule forms salt bridges to E410 in the  $\alpha$ 2<sup>S</sup> helix and to E345 in L1N<sup>S</sup> in the outer and inner molecules of the opposite dimer, respectively. In the hydrophobic core of interface 3, L402 and F403 in L2<sup>S</sup> of outer molecules interact with F493, F496, L655 and T651 of outer molecules in the neighbouring dimer (Fig. 1 and Extended Data Fig. 3a). Mutation of F403 and of E410 yielded predominantly dimeric protein and compromised liposome binding as well as liposome-stimulated GTPase activity (Fig. 2a–c). In dynamin 2, the F403A mutation substantially interfered with CME, as monitored by transferrin internalization (Fig. 2d). The effect of E410A on CME was less pronounced (Fig. 2d), as the structural defect may in part be compensated by the second salt bridge that R399 forms to E345 (Fig. 1, Extended Data Fig. 3a).

L1N<sup>S</sup> of an inner stalk also interacts with  $\alpha$ 1C<sup>S</sup> of an outer stalk. Accordingly, mutation of N-terminal (S347A/G348A/D349A), central (Q350A/V351A/D352A/T353A) or C-terminal (L354A/E355A/L356A/S357A) residue stretches in L1N<sup>S</sup> interfered with tetramerization (Fig. 2a). The central and C-terminal, but not the N-terminal mutations compromised liposome binding and assembly-stimulated GTPase activity (Fig. 2b, c). The Q350A/V351A/D352A/T353A mutant showed a reduced ability to sustain CME of transferrin, whereas the L354A/E355A/L356A/S357A mutant displayed a dominant-negative effect in transferrin uptake assays (Fig. 2d). The Charcot-Marie-Tooth neuropathy-related mutation G358R (refs 16, 17) is located in the C terminus of L1N<sup>S</sup> (Extended Data Fig. 3c). This mutation led to a dimeric mutant that did not bind to liposomes (Fig. 2a–c). Likewise, it exhibited a dominant-negative effect on CME (refs 16 and 17, and Fig. 2d). The bulky arginine side chain probably interferes with the proper binding conformation of L1N<sup>S</sup>. Interestingly, the mutants L354A/E355A/L356A/S357A, G358R and



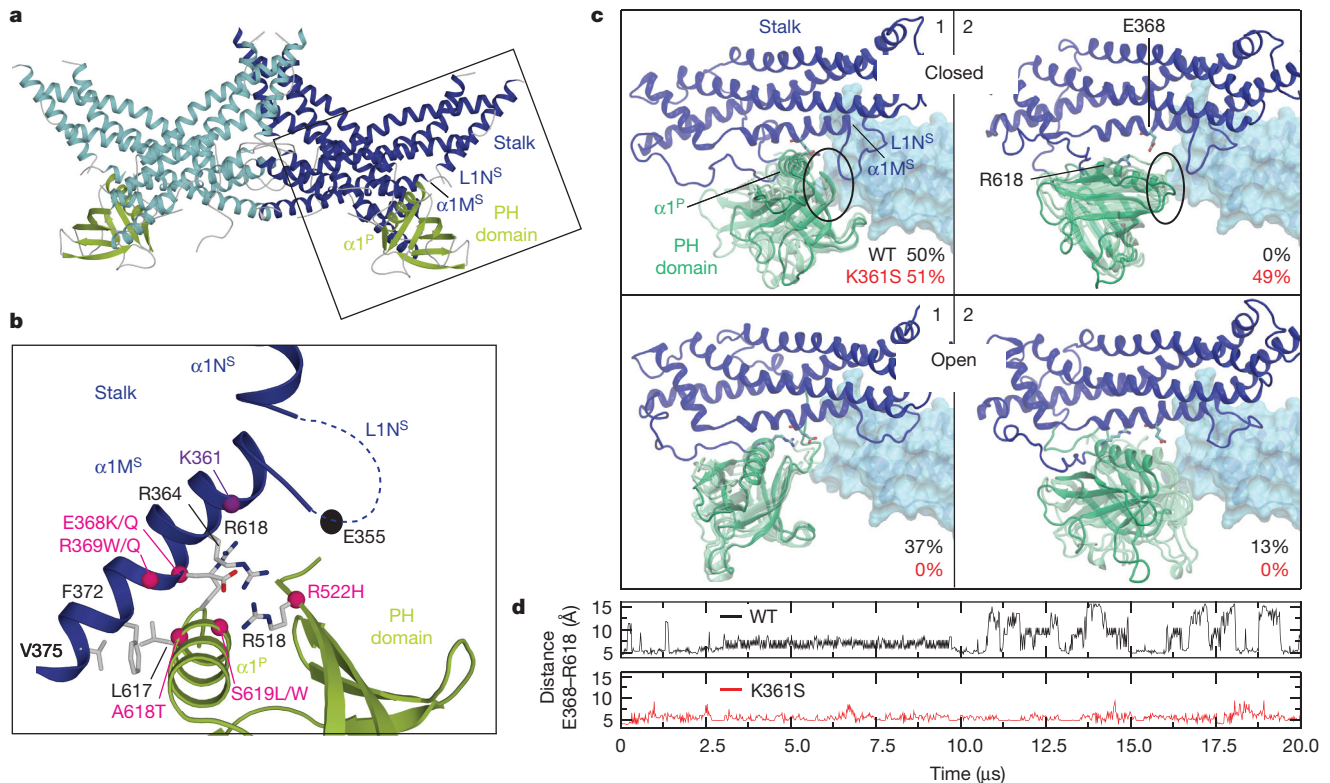
**Figure 2 | Interface 3 is crucial for assembly and function of dynamin.**

**a**, Sedimentation velocity experiments for dynamin 3 and the indicated mutants. The following molecular masses were obtained for singly sedimenting species: Q350A/V351A/D352A/T353A, 162 kDa; L354A/E355A/L356A/S357A, 164 kDa; G358R, 167 kDa. The molecular weight of the dynamin 3 construct is 86 kDa. S, Svedberg; AU, absorbance units. **b**, Liposome co-sedimentation assays for dynamin 3 and the indicated mutants. S, supernatant; P, pellet fraction; WT, wild type. **c**, The observed rate of GTPase activity for dynamin 3 and the indicated mutants in the absence or presence of liposomes. The average of two independent measurements is shown, with deviations ranging from 1% to 13%. **d**, The capacity of dynamin 2 mutants to reconstitute defective CME in HeLa cells depleted of endogenous dynamin 2, as monitored by fluorescent transferrin uptake. Data shown represent mean  $\pm$  s.e.m., the number of independent experiments is indicated in the bar. Sequence QIDT (amino acids 350–353) in dynamin 2 corresponds to QVDT in dynamin 3. siRNA, short interfering RNA. Raw data for **b** is available in the Supplementary Information.

F403A were still recruited to clathrin-coated pits; these pits, however, remained stable at the membrane surface (Extended Data Fig. 3d). Thus, the function of dynamin at clathrin-coated pits, but not its recruitment, depends on an intact interface 3.

The dimers in the dynamin tetramer are asymmetric concerning the PH domain and the orientation of the G domain and the BSE (Fig. 1). Compared to an outer molecule, G domains of inner molecules are tilted by approximately 40° around hinge 1 between the BSE and the stalk (Extended Data Fig. 4a). The PH domains of the outer molecules bind to a conserved surface of the stalk (Fig. 3a, b), a similar site as in dimeric dynamin 1 (ref. 2) (Extended Data Fig. 4b). The assignment of the visible PH domains to the outer molecules is unambiguous (Extended Data Fig. 4c). The PH domains of the inner molecules were not resolved in the electron density (Extended Data Fig. 4d). Modelling of the inner PH domains to positions equivalent to those observed for the outer molecules leads to clashes with the outer stalks (Fig. 3c and Extended Data Fig. 4e). Apparently, the PH domains have to be released from their autoinhibitory site for oligomerization to proceed. In keeping with this assumption, a dynamin 3 variant lacking the PH domain assembled in the absence of membranes and the presence of nucleotides more efficiently into regular oligomers than did wild-type dynamin 3 (Extended Data Fig. 5a–c). Dynamin 3 tubulated liposomes





**Figure 3 | Coupling of autoinhibition and oligomerization.** **a**, Stalks and PH domains in the dynamin tetramer as seen in the crystal. The box defines the view displayed in **c**. **b**, Close-up view of the PH-domain-stalk interface from **a**. Mutations in dynamin 3 implicated in centronuclear myopathy are indicated as pink balls; K361 and E355 as purple and black balls, respectively. **c**, Markov state models were constructed from MD simulation data including the stalk and PH domain. For each metastable PH domain conformation, three

on its own and did not need a specific membrane curvature for binding. At physiological salt concentrations, dynamin 3 efficiently bound to and tubulated unfiltered Folch liposomes (Extended Data Fig. 5d, e). This is in line with the presence of a tyrosine in position 596, which has been suggested to serve as a determinant for curvature generation versus curvature sensing<sup>18</sup>. When expressed in mammalian cells, dynamin 2( $\Delta$ PH) formed large, presumably cytosolic aggregates that failed to co-localize with clathrin and interfered with transferrin uptake in a dominant-negative fashion (Extended Data Fig. 5f, g), as has been shown previously for dynamin 1 (ref. 19). These results indicate that the PH domain has important functions in oligomerization and membrane binding.

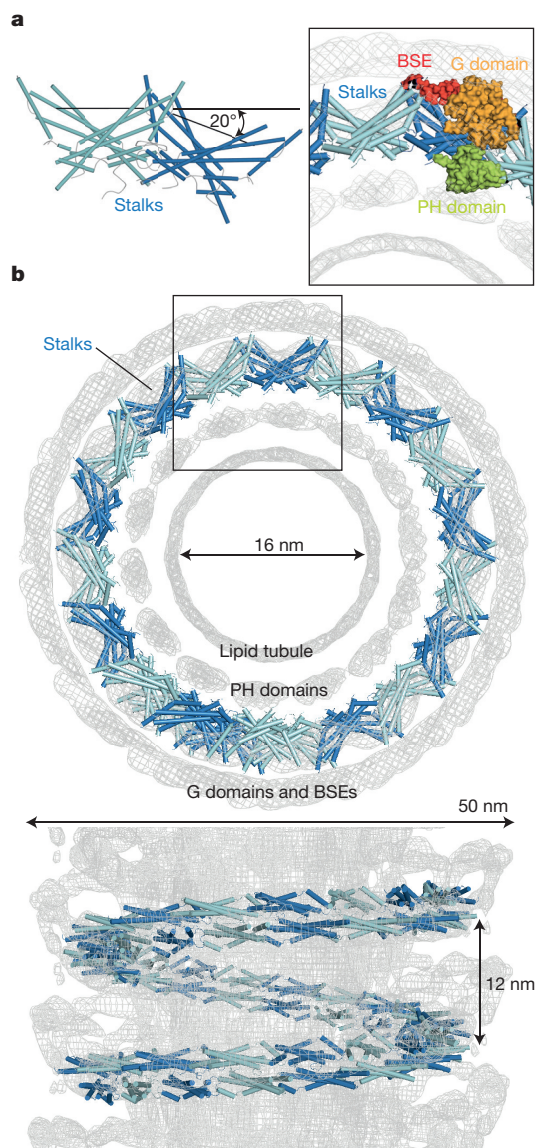
To investigate this dual function of the PH domain, we inserted the mutations R364S, R518H, R518D and E355A into the interface between the PH domain and stalk (Extended Data Fig. 6). Similar to nearby interface-3 mutants, most of these mutations impeded assembly, liposome binding, liposome-stimulated GTPase activity and transferrin uptake. In contrast, the mutation R518D enhanced oligomerization and GTP hydrolysis, as previously described for mutants in the PH-domain-stalk interface<sup>2,20</sup>.

Molecular dynamics (MD) simulations were carried out and analysed by Markov models<sup>21–23</sup> to characterize the dynamics of the PH-domain-stalk interface and its interplay with interface 3 (Fig. 3 and Extended Data Fig. 7). The simulations showed that E355 and K361, together with R518 and R364, are part of a network of polar interactions (Extended Data Fig. 7a) that can rapidly interconvert leading to three distinct binding modes (Fig. 3c). The preferred binding interaction of the PH domain with the stalk was the autoinhibitory ‘closed’ conformation found in our crystals. In two other ‘open’ conformations, the PH domain was shifted along the stalk to a position

representative structures are shown. Clashes between the PH domain and assembling stalk (light-blue surface) are indicated by black ovals. Percentages indicate the occurrence of each metastable state for wild-type dynamin 3 (black) and the mutant K361S (red). Open/closed: position of the PH domain allows/inhibits oligomerization. **d**, Example charged trajectories. The conformational dynamics are projected onto the Glu368–Arg618 distance, which illustrates opening and closure of the PH domain at the autoinhibited site.

where it did not interfere with oligomerization, indicating a dynamic equilibrium of oligomerization-permissive and non-permissive binding modes. The mutation K361S resulted in the appearance of a fourth, highly populated conformation that was also autoinhibitory for oligomerization, whereas no oligomerization-permissive binding modes were detected (Fig. 3c, d). Further MD simulations and Markov models of the stalk with a dissociated PH domain indicated that the mutant K361S predominantly stabilizes the loop L1N<sup>S</sup> in a conformation that is not adopted in the wild type (Extended Data Fig. 7c). Together, these results may explain the reduced oligomerization capacity of K361S, which is dimeric in solution (Extended Data Fig. 6c). A set of highly conserved charged residues including K361 apparently regulates both the autoinhibitory interaction with the PH domain and interactions with L1N<sup>S</sup>, thereby tightly coupling autoinhibition and oligomerization.

Comparison of the dynamin tetramer with the filament-like arrangements observed in the crystal structures of dynamin 1 (refs 2, 3) shows that the tetramer is bent, such that the angle between the outer stalks is changed by 20° (Figs 1 and 4a). We constructed a dynamin oligomer by stepwise addition of tetramers to the free ends of the growing dynamin assembly, using the geometry of interface 3 to connect the tetramers. This led to a right-handed helix (Fig. 4b), closely matching the dimensions of the dynamin 1 helix in the non-constricted, nucleotide-free state<sup>24</sup>. These observations indicate that formation of a right-handed dynamin helix at the surface of a tubular membrane is an intrinsic feature of stalk assembly via interface 3. The bent shape of the tetramer appears to dictate the curvature of a membrane tubule around which dynamin preferentially oligomerizes<sup>24,25</sup>. Constriction of membrane tubules to inner diameters smaller than 16 nm requires active GTP turnover and the associated G domain interactions across helical turns.



**Figure 4 | Assembly of the stalks leads to a right-handed dynamin helix.** **a**, Bent architecture of the dynamin 3 tetramer. Only the stalk helices are shown as cylinders, first dimer in light blue, second dimer in dark blue. **b**, Assembly of dynamin 3 tetramers using the geometry of interface 3 leads to a right-handed helix which fits an EM map of the non-constricted dynamin 1 helix<sup>24</sup> (shown in mesh representation). For clarity, only the stalks are displayed. In the inset, the G domain and BSE of an inner molecule and the PH domain of the adjacent outer molecule are shown in surface representation.

Comparison of our structure with a recent cryo-EM model of a super-constricted dynamin helix<sup>26</sup> (Extended Data Fig. 2c) suggests that constriction of the dynamin helix is driven by conformational changes in the stalk interfaces.

The stalks of our helix model fit well into a cryo-EM map of nucleotide-free dynamin 1 assembled around a lipid tubule<sup>24</sup> (Fig. 4b), but the PH domains and the G domains protrude from the electron density (Fig. 4b inset and Extended Data Fig. 8a). Apparently, the PH domains are shed from the autoinhibitory stalk interface to bind the membrane tubule, whereas the G domains move upwards from their positions. To explain the assembly of dynamin, we propose an equilibrium between PH domains bound (as seen for the outer molecules) and unbound (as for the inner molecules) to their stalks. In the cytosol, this equilibrium lies to the autoinhibited tetramer to prevent untimely oligomerization. Centronuclear-myopathy-related mutations in the interface between the stalk and PH domain (Fig. 3b and Extended Data Fig. 9) shift the

equilibrium towards the oligomerized state, thereby leading to disease. Upon dynamin recruitment by accessory proteins to endocytic sites, the equilibrium is driven towards the assembly-competent conformation. This hypothesis is supported by studies showing that, *in vivo*, dynamin helices are built by incorporation of dimer or tetramer units rather than larger preformed dynamin assemblies<sup>27,28</sup>. Further interactions, which may influence the assembly equilibrium, occur between the BSE and stalk, or the G domain and PH domain of adjacent dimers (Extended Data Fig. 8b–d). In this view, the effect of the disease-relevant mutation R465W may be explained. In the tetramer, R465 of an outer molecule is in close vicinity to the inner BSE of an adjacent dimer and a tryptophan at this position is likely to modify this interaction resulting in enhanced oligomerization.

A striking feature of dynamin assembly is the multitude of interactions in all four molecules of the tetramer (Extended Data Fig. 3a). Our results indicate that these contacts are not necessarily static, but are characterized by a dynamic equilibrium of different binding conformations. The formation of new interactions during assembly is compensated for by the release of autoinhibitory contacts in the dynamin tetramer. Such an assembly mode that involves many low-affinity interaction sites facilitates reversibility and allows regulation, for example through nucleotide binding, hydrolysis or phosphorylation<sup>29</sup>. It is the basis for the particular interaction mode of the semi-solid dynamin polymer with its protein and membrane environment, which has been previously identified in other CME proteins and has been coined as ‘matricity’<sup>30</sup>.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 18 December 2014; accepted 3 August 2015.

Published online 24 August 2015.

1. Ferguson, S. M. & De Camilli, P. Dynamin, a membrane-remodelling GTPase. *Nature Rev. Mol. Cell Biol.* **13**, 75–88 (2012).
2. Faelber, K. *et al.* Crystal structure of nucleotide-free dynamin. *Nature* **477**, 556–560 (2011).
3. Ford, M. G., Jenni, S. & Nunnari, J. The crystal structure of dynamin. *Nature* **477**, 561–566 (2011).
4. Cowling, B. S., Toussaint, A., Muller, J. & Laporte, J. Defective membrane remodeling in neuromuscular diseases: insights from animal models. *PLoS Genet.* **8**, e1002595 (2012).
5. Durieux, A. C., Prudhon, B., Guicheney, P. & Bitoun, M. Dynamin 2 and human diseases. *J. Mol. Med.* **88**, 339–350 (2010).
6. Daumke, O., Roux, A. & Haucke, V. BAR domain scaffolds in dynamin-mediated membrane fission. *Cell* **156**, 882–892 (2014).
7. Hinshaw, J. E. & Schmid, S. L. Dynamin self-assembles into rings suggesting a mechanism for coated vesicle budding. *Nature* **374**, 190–192 (1995).
8. Takei, K., McPherson, P. S., Schmid, S. L. & De Camilli, P. Tubular membrane invaginations coated by dynamin rings are induced by GTP-γS in nerve terminals. *Nature* **374**, 186–190 (1995).
9. Chappie, J. S. *et al.* A pseudoatomic model of the dynamin polymer identifies a hydrolysis-dependent powerstroke. *Cell* **147**, 209–222 (2011).
10. Faelber, K. *et al.* Structural insights into dynamin-mediated membrane fission. *Structure* **20**, 1621–1628 (2012).
11. Ramachandran, R. *et al.* The dynamin middle domain is critical for tetramerization and higher-order self-assembly. *EMBO J.* **26**, 559–566 (2007).
12. Fröhlich, C. *et al.* Structural insights into oligomerization and mitochondrial remodeling of dynamin 1-like protein. *EMBO J.* **32**, 1280–1292 (2013).
13. Gao, S. *et al.* Structure of myxovirus resistance protein a reveals intra- and intermolecular domain interactions required for the antiviral function. *Immunity* **35**, 514–525 (2011).
14. Gao, S. *et al.* Structural basis of oligomerization in the stalk region of dynamin-like MxA. *Nature* **465**, 502–506 (2010).
15. Srinivasan, S., Mattila, J. P. & Schmid, S. L. Intrapolypeptide Interactions between the GTPase Effector Domain (GED) and the GTPase Domain Form the Bundle Signaling Element in Dynamin Dimers. *Biochemistry* **53**, 5724–5726 (2014).
16. Koutsopoulos, O. S. *et al.* Mild functional differences of dynamin 2 mutations associated to centronuclear myopathy and Charcot-Marie Tooth peripheral neuropathy. *PLoS ONE* **6**, e27498 (2011).
17. Sidiropoulos, P. N. *et al.* Dynamin 2 mutations in Charcot-Marie-Tooth neuropathy highlight the importance of clathrin-mediated endocytosis in myelination. *Brain* **135**, 1395–1411 (2012).
18. Liu, Y. W. *et al.* Differential curvature sensing and generating activities of dynamin isoforms provide opportunities for tissue-specific regulation. *Proc. Natl Acad. Sci. USA* **108**, E234–E242 (2011).

19. Vallis, Y. *et al.* Importance of the pleckstrin homology domain of dynamin in clathrin-mediated endocytosis. *Curr. Biol.* **9**, 257–263 (1999).
20. Kenniston, J. A. & Lemmon, M. A. Dynamin GTPase regulation is altered by PH domain mutations found in centronuclear myopathy patients. *EMBO J.* **29**, 3054–3067 (2010).
21. Bowman, G. R., Pande, V. S. & Noé, F. (eds) *An Introduction to Markov State Models and Their Application to Long Timescale Molecular Simulation*. (Springer, 2014).
22. Schütte, C. & Sarich, M. Metastability and Markov models in molecular dynamics: modeling, analysis, algorithmic approaches. In *Courant Lecture Notes* Vol. 24 (American Mathematical Society, 2013).
23. Kohlhoff, K. J. *et al.* Cloud-based simulations on Google Exacycle reveal ligand modulation of GPCR activation pathways. *Nature Chem.* **6**, 15–21 (2014).
24. Chen, Y. J., Zhang, P., Egelman, E. H. & Hinshaw, J. E. The stalk region of dynamin drives the constriction of dynamin tubes. *Nature Struct. Mol. Biol.* **11**, 574–575 (2004).
25. Roux, A. *et al.* Membrane curvature controls dynamin polymerization. *Proc. Natl Acad. Sci. USA* **107**, 4141–4146 (2010).
26. Sundborger, A. C. *et al.* A dynamin mutant defines a super-constricted pre-fission state. *Cell Rep.* **8**, 734–742 (2014).
27. Cocucci, E., Gaudin, R. & Kirchhausen, T. Dynamin recruitment and membrane scission at the neck of a clathrin-coated pit. *Mol. Biol. Cell* **25**, 3595–3609 (2014).
28. Grassart, A. *et al.* Actin and dynamin2 dynamics and interplay during clathrin-mediated endocytosis. *J. Cell Biol.* **205**, 721–735 (2014).
29. Graham, M. E., O'Callaghan, D. W., McMahon, H. T. & Burgoyne, R. D. Dynamin-dependent and dynamin-independent processes contribute to the regulation of single vesicle release kinetics and quantal size. *Proc. Natl Acad. Sci. USA* **99**, 7124–7129 (2002).
30. Schmid, E. M. & McMahon, H. T. Integrating molecular and network biology to decode endocytosis. *Nature* **448**, 883–888 (2007).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** This project was supported by grants from the Deutsche Forschungsgemeinschaft (MA1081/8-2 to D.J.M.; SFB740/D7 and SFB958/A04 to F.N.; SFB740/C8 and SFB958/A7 to V.H.; SFB740/C7 and SFB958/A12 to O.D.; and ES410/2-1 to S.E.), an ERC consolidator grant (ERC-2013-CoG-616024 to O.D.), an ERC starting grant (pcCell to F.N.) and a grant from the Einstein Foundation Berlin (SOoPiC to N.P.). T.F.R. acknowledges partial financial support by the Cluster of Excellence REBIRTH (DFG EXC 62/1). We thank B. Purfürst for help with electron microscopy; S. Hertel, L. Litz, P. Straub and S. Wohlgemuth for experimental assistance; and the staff at beamlines X06SA (PXI) and X06DA (PXIII) at the Swiss Light Source (Villigen, Switzerland) for help during data collection. We thank Y.-W. Liu for discussions, and A. Wittinghofer for his support and discussions in the initial stages of the project.

**Author Contributions** T.F.R. grew the crystals and collected data; K.F. solved the structure; T.F.R., K.F. and S.E. refined the structure; T.F.R. and R.A. purified protein for crystallization and monomeric dynamin; K.F. and J.S. purified all other proteins, performed liposome co-sedimentation, EM and GTPase assays; U.C. performed and analysed analytical ultracentrifugation experiments; Y.P. and K.K. performed transferrin uptake assays; N.P. and F.N. conducted and analysed molecular modelling and molecular dynamics simulations. N.P. and Y.P. contributed equally to this work. T.F.R., K.F., F.N., V.H., O.D. and S.E. interpreted structural data. T.F.R., K.F., Y.P., N.P., U.C., D.J.M., F.N., V.H., O.D. and S.E. designed the research. T.F.R., K.F., F.N., O.D. and S.E. wrote the manuscript.

**Author Information** The atomic coordinates and structure factors of human dynamin 3 have been deposited in the Protein Data Bank (PDB) with accession number 5A3F. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to K.F. ([katja.faelber@mdc-berlin.de](mailto:katja.faelber@mdc-berlin.de)), O.D. ([oliver.daumke@mdc-berlin.de](mailto:oliver.daumke@mdc-berlin.de)), or S.E. ([Eschenburg.Susanne@mh-hannover.de](mailto:Eschenburg.Susanne@mh-hannover.de)).



## METHODS

**Protein expression and purification.** Human dynamin 3 (splice form abb<sup>31</sup>, residues 1–754) and indicated mutants of this construct were expressed from the pProEx-HTb vector (Invitrogen) as N-terminal His<sub>6</sub>-tag fusion followed by a tobacco etch virus (TEV) cleavage site. The crystallized construct contained the K361S mutation. Proteins were produced in *Escherichia coli* host strain BL21(DE3), and expression was induced by addition of 0.1 mM isopropyl- $\beta$ -D-thiogalactopyranoside. Cells were grown over night at 20 °C in terrific broth medium. The following procedure was used for purification of dynamin 3(K361S) for crystallization. Cells were resuspended in buffer A300 (50 mM HEPES-NaOH (pH 7.5), 300 mM NaCl, 15 mM imidazole and 2 mM MgCl<sub>2</sub>) including 1 mM phenylmethylsulfonyl fluoride and 0.1% v/v NP-40, and disrupted by sonification. Cleared lysates (30,000g, 1 h, 4 °C) were applied to a Ni<sup>2+</sup>-NTA column (Qiagen). The column was sequentially washed with buffer A300 and with buffer A100 (100 mM NaCl). Protein was eluted with buffer A100 containing an additional 285 mM imidazole. Fractions containing human dynamin 3 were pooled and diluted with an equal volume of 50 mM HEPES-NaOH (pH 7.5). The diluted protein was loaded onto a HiLoad SuperQ anion exchange column (GE Healthcare) equilibrated with buffer B50 (in which 50 refers to the NaCl concentration) containing 50 mM HEPES-NaOH (pH 7.5), 50 mM NaCl, 2 mM DTE and 1 mM MgCl<sub>2</sub>. After washing with buffer B50, bound proteins were eluted with a linear gradient from 50 to 500 mM NaCl. Fractions containing human dynamin 3 were pooled, 1 mg TEV per 10 mg dynamin 3 was added, and the protein incubated on ice for 4 h. The solution was concentrated using 50 kDa molecular weight cut-off concentrators (Amicon) and applied onto a Superdex 200 gel filtration column (GE Healthcare) equilibrated with buffer B100. Fractions containing dynamin 3 were pooled, concentrated and flash-frozen in liquid nitrogen.

Wild-type and mutant dynamin 3 used for biochemical and biophysical assays were expressed in *E. coli* Rosetta2-BL21-DE3 in autoinduction medium (Novagen) and purified using a Co<sup>2+</sup>-Talon column, followed by overnight TEV cleavage (4 °C, 30  $\mu$ g per 1 mg fusion protein), dilution/concentration in concentrators for imidazole removal and a second Co<sup>2+</sup>-Talon column run for His<sub>6</sub>-TEV and uncleaved His<sub>6</sub>-dynamin capture. Finally, the peak fractions from a Superdex 200 gel filtration containing dynamin were pooled, concentrated to maximal 20 mg ml<sup>-1</sup> and flash frozen in liquid nitrogen. The purification buffer contained 20 mM HEPES-NaOH, pH 7.8, 500 mM NaCl, and 2 mM MgCl<sub>2</sub> (plus 100 mM imidazole for elution, plus 2.8 mM  $\beta$ -mercaptoethanol during TEV cleavage). The purified protein was nucleotide-free, as confirmed by high-performance liquid chromatography analysis (see below for details).

**Crystallization and structure determination.** Crystallization trials by the sitting-drop vapour-diffusion method were performed at 4 °C using a mosquito LCP pipetting robot (TTP Labtech) and Rock Imager storage system (Formulatrix). Human dynamin 3 in 150 nl volumes at a concentration of 20 mg ml<sup>-1</sup> was mixed with an equal volume of reservoir solution from commercially available preformulated screens (Qiagen). On a preparative scale, 2  $\mu$ l of protein solution was mixed with 2  $\mu$ l of reservoir solution containing 100 mM MES-NaOH (pH 6.5) and 15% 2-methyl-2,4-pentanediol. Crystals appeared after three to five days and reached final dimensions of up to 0.5 mm  $\times$  0.3 mm  $\times$  0.3 mm. Crystals were cryoprotected by immersion in reservoir solution added with increasing amounts of ethylene glycol with the final solution containing 17% v/v ethylene glycol. Cryoprotected crystals were flash-cooled in liquid nitrogen. Data were recorded at beamline PXI-X06SA at the Swiss Light Source (Villigen, Switzerland). Native data from a single crystal was processed and scaled using the program package XDS<sup>32</sup>. The structure was solved by molecular replacement with Phaser<sup>33</sup> using the structure of the nucleotide-free rat dynamin 1 G domain (2AKA), the stalk of human dynamin (3SNH) and the human PH domain (1DYN) as search models. The model was built using Coot<sup>34</sup> and iteratively refined using Phenix<sup>35</sup> with noncrystallographic symmetry (NCS) between the outer and the inner molecules, respectively, with reference model restraints against an artificial dynamin construct composed of the high resolution search model domains, and with one translation-libration-screw (TLS) group per domain.

Owing to weak electron density, all residues of the G domains of the inner molecules were chopped at the C $\beta$  atoms, and the whole domains were refined as rigid bodies. In the final model, the outer molecules have disordered regions in the L1N<sup>S</sup> loop, the L1<sup>SP</sup> and L2<sup>SP</sup> loops and the L5<sup>P</sup> loop, and the inner molecules in the hinge 1 region and the L2<sup>S</sup> loop. Furthermore, the complete PH domains of the inner molecules are not resolved in the electron density. The structure was refined to  $R_{\text{work}}/R_{\text{free}}$  of 23.2%/27.8%. Of all residues, 94.5% are in the most favoured regions of the Ramachandran plot and 0.6% (15 out of 2,500) of residues are in the disallowed regions, as analysed with Molprobity<sup>36</sup>. Figures were prepared with PyMOL<sup>37</sup>. Domain superpositions were performed with lsqkab<sup>38</sup>. Sequences were aligned using Clustal W<sup>39</sup> and adjusted by hand. The model of the

right-handed dynamin 3 helix was fitted manually into the EM map using PyMOL<sup>37</sup> and Chimera<sup>40</sup>.

**Analytical ultracentrifugation.** Sedimentation velocity experiments were carried out in a ProteomeLab XL-I analytical ultracentrifuge (Beckmann Coulter) at 35,000 r.p.m. and 20 °C using an An-50 Ti rotor. Concentration profiles were measured using the manufacturer's data acquisition software ProteomeLab XL-I Version 6.0 (Firmware 5.7) with the absorption scanning optics at 280 nm. Sedimentation velocity analysis was performed in a buffer containing 0.15 M NaCl, 50 mM HEPES-NaOH (pH 7.5) in 3 or 12 mm standard double sector centrepieces filled with a 100  $\mu$ l or 400  $\mu$ l sample, respectively. For data analysis, a model for diffusion-deconvoluted differential sedimentation coefficient distributions (continuous  $c(s)$  distributions) implemented in the program SEDFIT<sup>41</sup> was used. For proteins sedimenting as a single species, molecular masses were obtained from  $c(s)$  analysis as calculated from the  $s$ -value and diffusion broadening of the sedimenting boundary. Dynamin 3(K361S/R399A/ $\Delta$ PRD) mutant, analysed in a concentration range from 4 to 23  $\mu$ M, showed a single peak in  $c(s)$  distributions with a sedimentation coefficient slightly decreasing with increasing protein concentration (data not shown). Owing to hydrodynamic non-ideality, this is expected for a protein that does not change its oligomerization state with concentration<sup>42</sup>. Extrapolation to zero concentration yielded  $s_{20,w} = 6.4$  S and a molecular mass of 160 kDa was obtained from  $c(s)$  analysis. Since the molecular mass of the monomer as calculated from the amino acid composition is 86 kDa, this mutant forms dimers in solution. For comparison, all other mutants were analysed at a concentration of about 20  $\mu$ M. The following molecular masses were obtained from  $c(s)$  analyses of mutants that sedimented as a single species: Q350A/V351A/D352A/T353A, 162 kDa; L354A/E355A/L356A/S357A, 164 kDa; G358R, 167 kDa.

Partial specific volume, buffer density and viscosity were calculated from amino acid and buffer composition, respectively, by the program SEDNTERP<sup>43</sup> and were used to correct experimental  $s$ -values to  $s_{20,w}$ . Figures were prepared using the program GUSI (<http://biophysics.swmed.edu/MBR/software.html>, provided by C. Brautigam).

**Liposome co-sedimentation assays.** Liposomes were prepared as previously described (<http://www.endocytosis.org>). Folch liposomes (total bovine brain lipids fraction I from Sigma) in 20 mM HEPES-NaOH (pH 7.5), 100 mM NaCl were extruded 13 times through a 0.1  $\mu$ m filter. The resulting 0.2 mg ml<sup>-1</sup> liposomes were incubated at room temperature with 4.0  $\mu$ M of the indicated dynamin 3 construct for 10 min in 40  $\mu$ l reaction volume, followed by a 213,000g spin for 10 min at 20 °C. The final reaction buffer contained 25 mM HEPES-NaOH pH 7.5, 140 mM NaCl, 2 mM MgCl<sub>2</sub> and 1 mM KCl.

**GTP hydrolysis assay.** GTPase activities of 1  $\mu$ M of the indicated dynamin constructs were determined at 37 °C in 25 mM HEPES-NaOH (pH 7.5), 130 mM NaCl, 2 mM MgCl<sub>2</sub> and 1 mM KCl, in the absence and presence of 0.1 mg ml<sup>-1</sup> 0.1- $\mu$ m filtered Folch liposomes, using saturating concentrations of GTP as substrate (1 mM for the basal and 3 mM for the stimulated reactions). Reactions were initiated by the addition of protein to the reaction. At different time points, reaction aliquots were 15-fold diluted and quickly transferred to liquid nitrogen. Nucleotides in the samples were separated via a reversed-phase Hypersil ODS-2 C18 column (250  $\times$  4 mm), with 100 mM potassium phosphate buffer (pH 6.5), 10 mM tetrabutylammonium bromide and 7.5% acetonitrile as running buffer. Denatured proteins were adsorbed on a C18 guard column. Nucleotides were detected by absorption at 254 nm and quantified by integration of the corresponding peaks. Rates were derived from a linear fit to the initial reaction (<20% GTP hydrolysed).

**Electron microscopy.** For electron microscopic studies (Zeiss EM910), 2  $\mu$ M dynamin 3 in 25 mM HEPES-NaOH (pH 7.5), 60 or 150 mM NaCl, 2 mM MgCl<sub>2</sub>, 1 mM KCl and 1 mM guanosine-5'-[( $\beta$ , $\gamma$ )-methylene]triphosphate were incubated at room temperature for 20 h without liposomes or for 20 min with liposomes. The final concentration of unfiltered liposomes was 0.35 mg ml<sup>-1</sup>. Samples were spotted on carbon-coated copper grids (Plano GmbH) and negatively stained with 3% uranyl acetate.

**Transferrin uptake in HeLa cells.** HeLa cells (ATCC no. CCL2, identified by their short tandem repeat and isoenzyme profiles) were obtained from ATCC (American Type Culture Collection) and tested for mycoplasma contamination on a routine basis. Cells were cultured for 20–25 passages without further authentication before starting a fresh culture from the immediately ATCC-derived stock. HeLa cells were transfected with siRNA using oligofectamine (Invitrogen). The sequence of the siRNA used to target human dynamin 2 was 5'-GCAACUGA CCAACCACAUC-3' (nucleotides 849–867). After 24 h, cells were transfected with pEGFP-N1 (Clontech) or siRNA-resistant rat dynamin 2-pEGFP-N1 using lipofectamine 2000 (Invitrogen). 72 h after siRNA-transfection, cells were serum-starved for 1 h and incubated with 15  $\mu$ g ml<sup>-1</sup> transferrin-Alexa 647 (Molecular Probes, Invitrogen) for 10 min at 37 °C. On ice, cells were washed once with cold



PBS and 10 mM MgCl<sub>2</sub>, and once for 90 s with 0.1 M acetic acid (pH 5.3) and 200 mM NaCl to remove surface-bound transferrin. After two washes with cold PBS and 10 mM MgCl<sub>2</sub>, cells were detached from the culture dish by incubating for 5 min on ice with 0.1% Pronase E solution in PBS and 0.5 mM EDTA. Cells were resuspended in 1% bovine serum albumin (BSA) in PBS, pelleted at 300g for 5 min at 4 °C and then fixed in 4% paraformaldehyde, 4% sucrose in PBS for 15 min on ice and a further 20 min at room temperature. Cells were pelleted and resuspended in 1% BSA in PBS and analysed by flow cytometry using a BD FACScalibur. Transferrin fluorescence in GFP-positive cells was quantified and normalized to cells rescued with wild-type dynamin 2-eGFP. No statistical methods were used to predetermine sample size.

**Localization of dynamin 2-eGFP mutants and analysis of clathrin-coated pit dynamics.** HeLa cells depleted of endogenous dynamin 2 as described above were co-transfected with plasmids encoding eGFP or dynamin 2-eGFP and mRFP-clathrin-light-chain. 72 h after siRNA transfection, cells were analysed by total internal reflection fluorescence (TIRF) microscopy using a Nikon Eclipse Ti (Andor sCMOS camera, Okolab incubator, Nikon PerfectFocus autofocus system, 60× TIRF-objective, operated by open source ImageJ-based Micromanager software<sup>44</sup>). For live imaging, cells growing on glass coverslips were kept in Hank's balanced salt solution with 5% fetal bovine serum. From 180 s dual-colour TIRF recordings with a frame rate of 0.5 Hz, kymographs were created by selecting a line of pixels from an individual cell and depicting this line over the duration of 90 frames.

**Molecular dynamics simulation and modelling.** Molecular dynamics (MD) simulations of the stalk and the pleckstrin homology (PH) domain (residues 322 to 710) were carried out for the wild type and the K361S mutant, each using three different setups: (1) the crystal structure coordinates of chain C superimposed to chain B were taken as starting point; (2) starting from setup 1, the PH domain was moved 5 Å away from the stalk; and (3) the PH domain was absent.

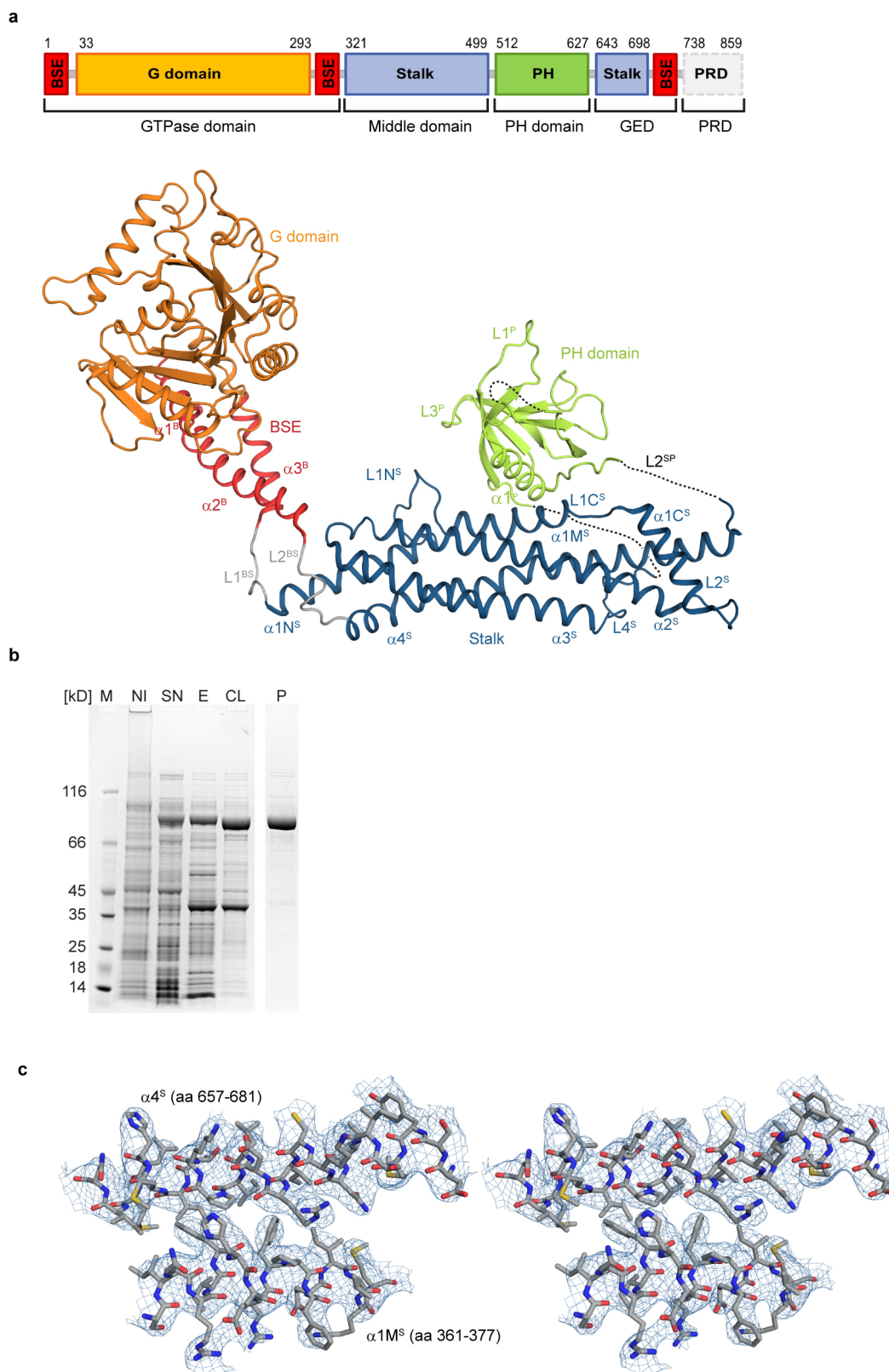
Setups 1 and 2 were used in order to study the conformational equilibrium of stalk-PH domain interactions. The loops joining the stalk to the PH domain (residues 495 to 511 and 628 to 640) were generated for each setup and mutant using VMD (visual molecular dynamics)<sup>45</sup> and were minimized and equilibrated separately. The aim of the setup 3 was to study the intrinsic conformational dynamics of the L1N<sup>S</sup> loop when the PH domain is dissociated from the stalk.

The coordinates of each setup and for each mutant were used to construct an all-atom molecular model and run MD simulations in explicit solvent with GROMACS<sup>46</sup> using the CHARMM27 force field<sup>47</sup>. For the setup and equilibration procedure, hydrogen atoms were added based on the heavy atom coordinates followed by an initial energy minimization. The protein was then solvated in a water box with a solvation layer of at least 10 Å, resulting in an overall system of between 70,000 and 80,000 atoms (depending on the initial structure). Na<sup>+</sup> and Cl<sup>-</sup> ions (100 mM) were added to buffer the system and obtain an overall neutral simulation cell. The solvated and ionized system was again minimized and equilibrated in the NVT (canonical) ensemble at 300 K with position constraints on the protein heavy atoms. A second equilibration was carried out in the NPT (isothermal-isobaric) ensemble, again with position constraints, followed by a 1 ns equilibration without constraints. The equilibrated coordinates and velocities were used as the starting point for twenty 100-ns production runs for each setup and with both wild type and mutant K361S, giving rise to a total of 12 microseconds of molecular dynamics data.

**Analysis with Markov state models.** The conformations of the L1N<sup>S</sup> loop and the stalk-PH domain patterns can be well characterized by their hydrogen bonding patterns within the loop or between stalk and PH domains. Here, 21 residue pairs, shown in Extended Data Fig. 7d, were selected that can form hydrogen bonds or salt bridges. The C $\alpha$  distances of these residue pairs were evaluated in order to obtain a low-dimensional representation of the respective configuration. These distances were used to build Markov state models<sup>21–23,48,49</sup> of setup 3 (L1N<sup>S</sup> loop)

and using both setups 1 and 2 for the PH-domain-stalk interactions using the EMMA program (<http://pyemma.org>)<sup>50</sup>. The microstates of the Markov state models were obtained by regular spatial clustering in the distance space<sup>48</sup>. The distance cutoff for the regular spatial clustering was chosen to obtain around 500 microstates. Using a cutoff of 10.75 Å for the wild type and a cutoff of 9 Å for the mutant K361S resulted in 550 and 584 microstates, respectively. The lag-time-dependent relaxation timescales, indicating approximate Markovianity<sup>51</sup> at lag times of 20 ns or larger, are shown in Extended Data Fig. 7. Reversible transition matrices were then estimated at a lag time of 20 ns. The microstates of each Markov state model were clustered into a set of three to four metastable states using the robust Perron cluster analysis<sup>52</sup>. At this resolution, the metastable states sampled by the different mutants can be clearly associated between wild type and mutant, as shown in Fig. 3c. The Markov model was used to generate random trajectories shown in Fig. 3d as described in ref. 23.

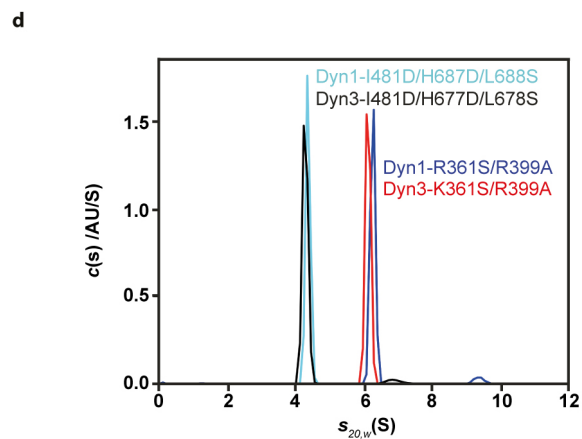
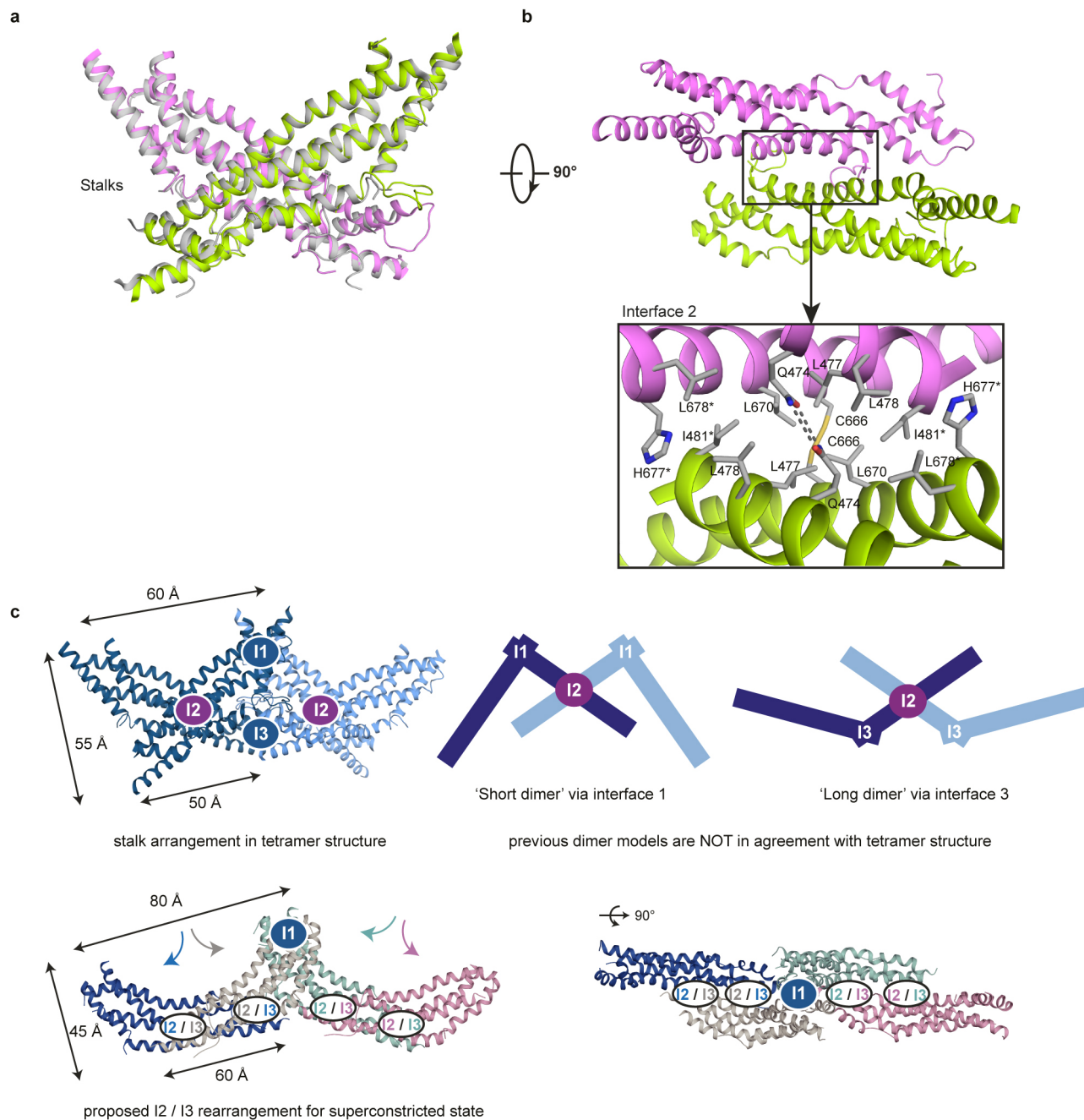
31. Cao, H., Garcia, F. & McNiven, M. A. Differential distribution of dynamin isoforms in mammalian cells. *Mol. Biol. Cell* **9**, 2595–2609 (1998).
32. Kabsch, W. XDS. *Acta Crystallogr. D* **66**, 125–132 (2010).
33. McCoy, A. J. et al. Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
34. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
35. Adams, P. D. et al. The Phenix software for automated determination of macromolecular structures. *Methods* **55**, 94–106 (2011).
36. Chen, V. B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66**, 12–21 (2010).
37. The PyMOL Molecular Graphics System. Version 1.7.0.1. (Schrödinger, LLC).
38. Kabsch, W. A solution for the best rotation to relate two sets of vectors. *Acta Crystallogr. A* **32**, 922–923 (1976).
39. Larkin, M. A. et al. Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947–2948 (2007).
40. Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
41. Schuck, P. Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. *Biophys. J.* **78**, 1606–1619 (2000).
42. Laue, T. M. & Stafford, W. F. III. Modern applications of analytical ultracentrifugation. *Annu. Rev. Biophys. Biomol. Struct.* **28**, 75–100 (1999).
43. Laue, M. T., Shah, B. D., Ridgeway, T. M. & Pelletier, S. L. in *Analytical Ultracentrifugation in Biochemistry and Polymer Science* (eds Harding, S. E. et al.) 90–125 (Royal Society of Chemistry, 1992).
44. Edelstein, A. et al. Computer control of microscopes using  $\mu$ Manager. *Curr. Protoc. Mol. Biol.* <http://dx.doi.org/10.1002/0471142727.mb1420592> (2010).
45. Humphrey, W., Dalke, A. & Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graph.* **14**, 33–38 (1996).
46. Lindahl, E., Hess, B. & van der Spoel, D. GROMACS 3.0: a package for molecular simulation and trajectory analysis. *J. Mol. Model.* **7**, 306–317 (2001).
47. Mackerell, A. D. Jr, Feig, M. & Brooks, C. L. III. Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J. Comput. Chem.* **25**, 1400–1415 (2004).
48. Prinz, J.-H. et al. Markov models of molecular kinetics: Generation and validation. *J. Chem. Phys.* **134**, 174105 (2011).
49. Stanley, N., Esteban-Martin, S. & De Fabritiis, G. Kinetic modulation of a disordered protein domain by phosphorylation. *Nature Commun.* **5**, 5272 (2014).
50. Senne, M. et al. EMMA: A Software Package for Markov Model Building and Analysis. *J. Chem. Theory Comput.* **8**, 2223–2238 (2012).
51. Swope, W. C., Pitera, J. W. & Suits, F. Describing protein folding kinetics by molecular dynamics simulations. 1. Theory. *J. Phys. Chem. B* **108**, 6571–6581 (2004).
52. Deuffhard, P. & Weber, M. Robust Perron cluster analysis in conformation dynamics. *Linear Algebra Appl.* **398**, 161–184 (2005).
53. Chappie, J. S. & Dyda, F. Building a fission machine—structural insights into dynamin assembly and activation. *J. Cell Sci.* **126**, 2773–2784 (2013).



### Extended Data Figure 1 | Characterization of the dynamin 3 construct.

**a**, Top: the domain structure of dynamin 3. The previously used sequence-derived domain nomenclature is shown below. Bottom: a dynamin 3 monomer colour-coded according to the domain architecture. **b**, SDS-PAGE representing a typical purification of dynamin 3. M, marker proteins; NI,

whole-cell-lysate non-induced; SN, supernatant of cleared lysate; E, elution peak of the Talon-Co<sup>2+</sup> column; CL, after cleavage with TEV protease; P, pool after gel filtration. **c**, Representative electron density map (stereo view). Two stalk helices are shown as stick models, the  $2F_o - F_c$  map is contoured at  $1.0\sigma$ . Raw data for **b** is available in Supplementary Information.



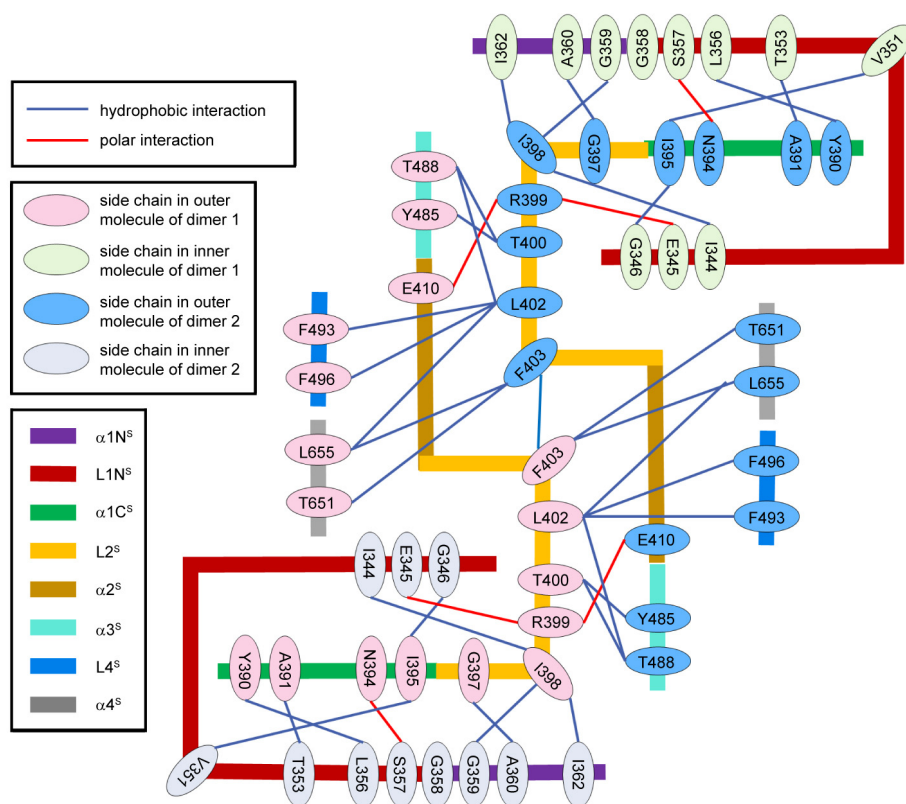
**Extended Data Figure 2 | Dimerization of dynamin 3.** **a**, Superposition of dynamin 1 (grey; PDB code: 3SNH) and dynamin 3 (magenta and green) dimers, colour-coded as in Fig. 1. The stalk arrangement in dynamin 3 is essentially the same as in dynamin 1. **b**, Interface 2 in dynamin 3. The view is rotated by 90° with respect to **a**. The zoom shows the side chains of residues involved in interface formation. Residues, whose mutation render dynamin 1 and dynamin 3 monomeric, are marked with an asterisk. **c**, Top: stalks of the dynamin 3 tetramer, as seen in the crystal structure (left). Dynamin dimers (dark and light blue) are formed via interface 2 (I2) and assemble into the tetramer via interfaces 1 and 3 (I1 and I3, respectively). In alternative

dimerization models (middle and right)<sup>52</sup>, dynamin monomers assemble via interface 1 (middle) or interface 3 (right) to form elongated dimers of different shapes. Bottom: arrangement of stalks of a dynamin 1 as fitted into a cryo-EM density map of a super-constricted dynamin 1 helix (PDB code: 4UUD)<sup>26</sup>.

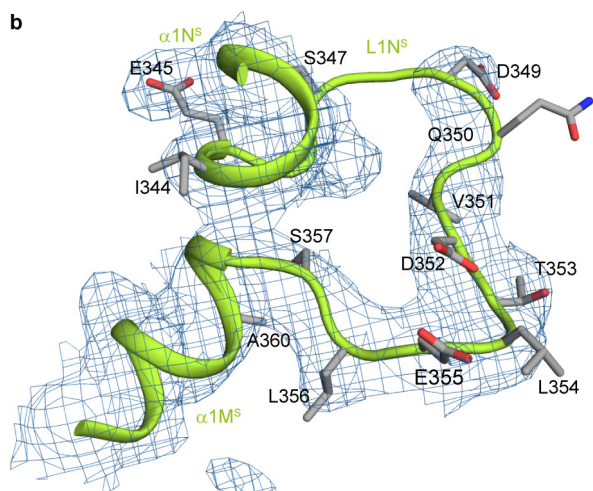
**d**, Oligomeric state of dimer interface mutants, as assayed by analytical ultracentrifugation at a protein concentration of 20 µM. The following molecular masses were obtained from *c(s)* analyses: dynamin 1(R361S/R399A) (176 kDa, dimeric) in dark blue; dynamin 1(I481D/H687D/L688S) (84 kDa, monomeric) in light blue; dynamin 3(K361S/R399A) (165 kDa, dimeric) in red, dynamin 3(I481D/H677D/L678S) (83 kDa, monomeric) in black.



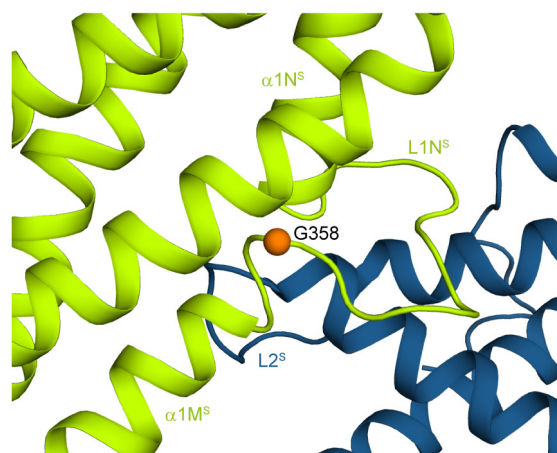
**a**



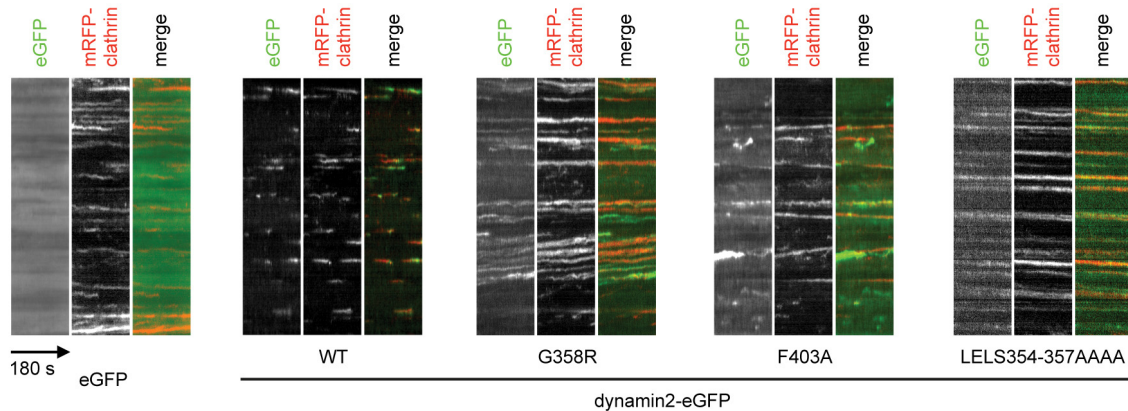
**b**



**C**

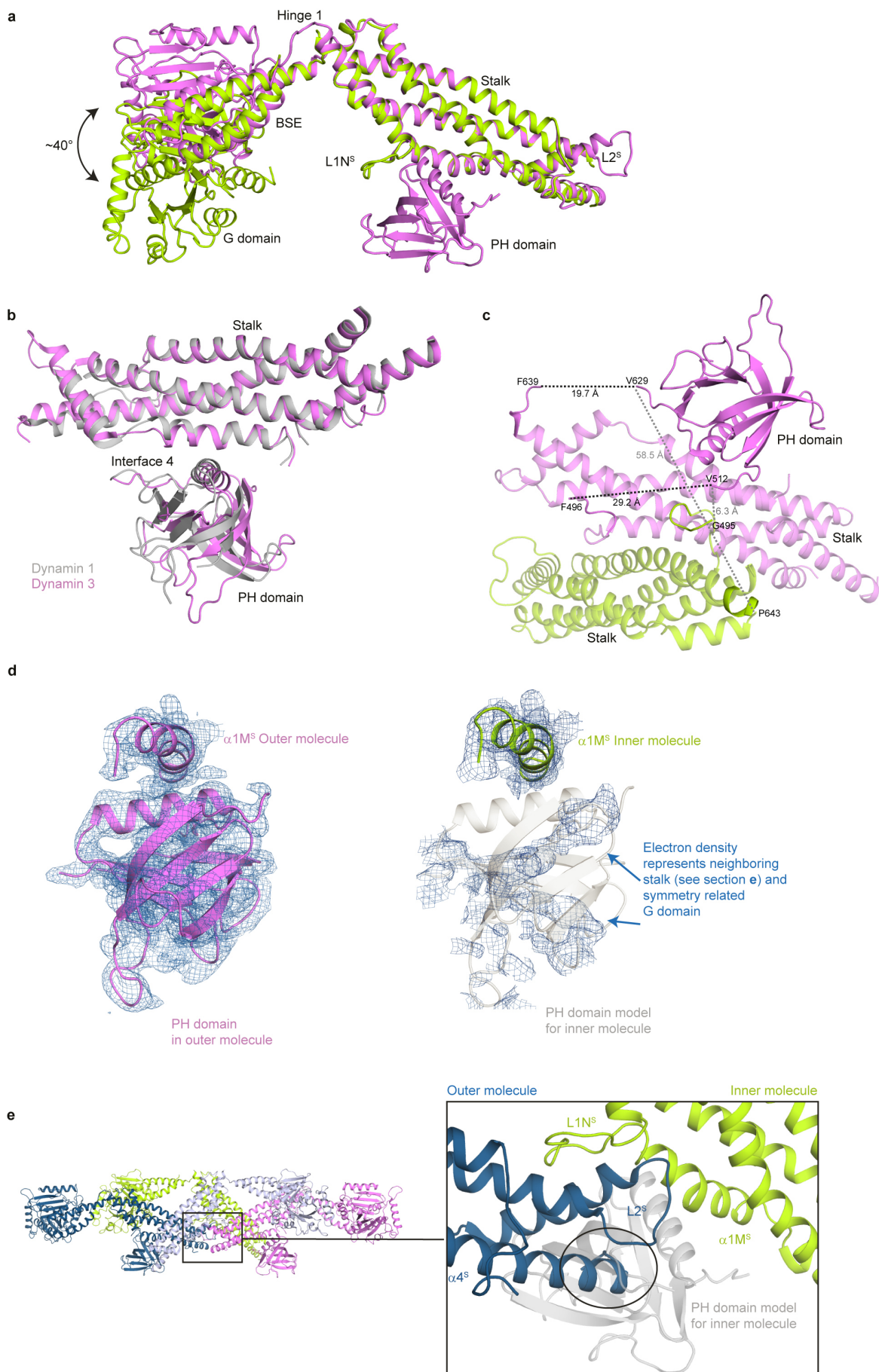


**d**



**Extended Data Figure 3 | Dynamin assembly via interface 3.** **a**, Schematic overview of the interactions in interface 3. **b**, Details of loop L1N<sup>S</sup>. The  $2F_o - F_c$  electron density is contoured at  $1.0\sigma$ . **c**, The Charcot-Marie-Tooth-related mutation G358R is located at the C-terminal end of loop L1N<sup>S</sup>. It probably disturbs the structural integrity of this loop and therefore might interfere with oligomerization. **d**, Clathrin-coated pit dynamics in HeLa cells expressing interface 3 mutants of dynamin 2. HeLa cells treated with dynamin 2 siRNA were co-transfected with plasmids encoding eGFP or siRNA-resistant dynamin

2-eGFP and mRFP-clathrin-light-chain, and live cells were imaged at 37 °C by TIRF microscopy. Shown are representative time-resolved line scans (kymographs) from at least ten time-lapse recordings of individual cells. Attenuated clathrin-coated pit dynamics upon depletion of endogenous dynamin 2 are only rescued by re-expression of wild-type but not mutant dynamin 2-eGFP. Note that the dynamin 2 mutants tested displayed a more diffuse subcellular distribution although they were still recruited to clathrin-coated pits.

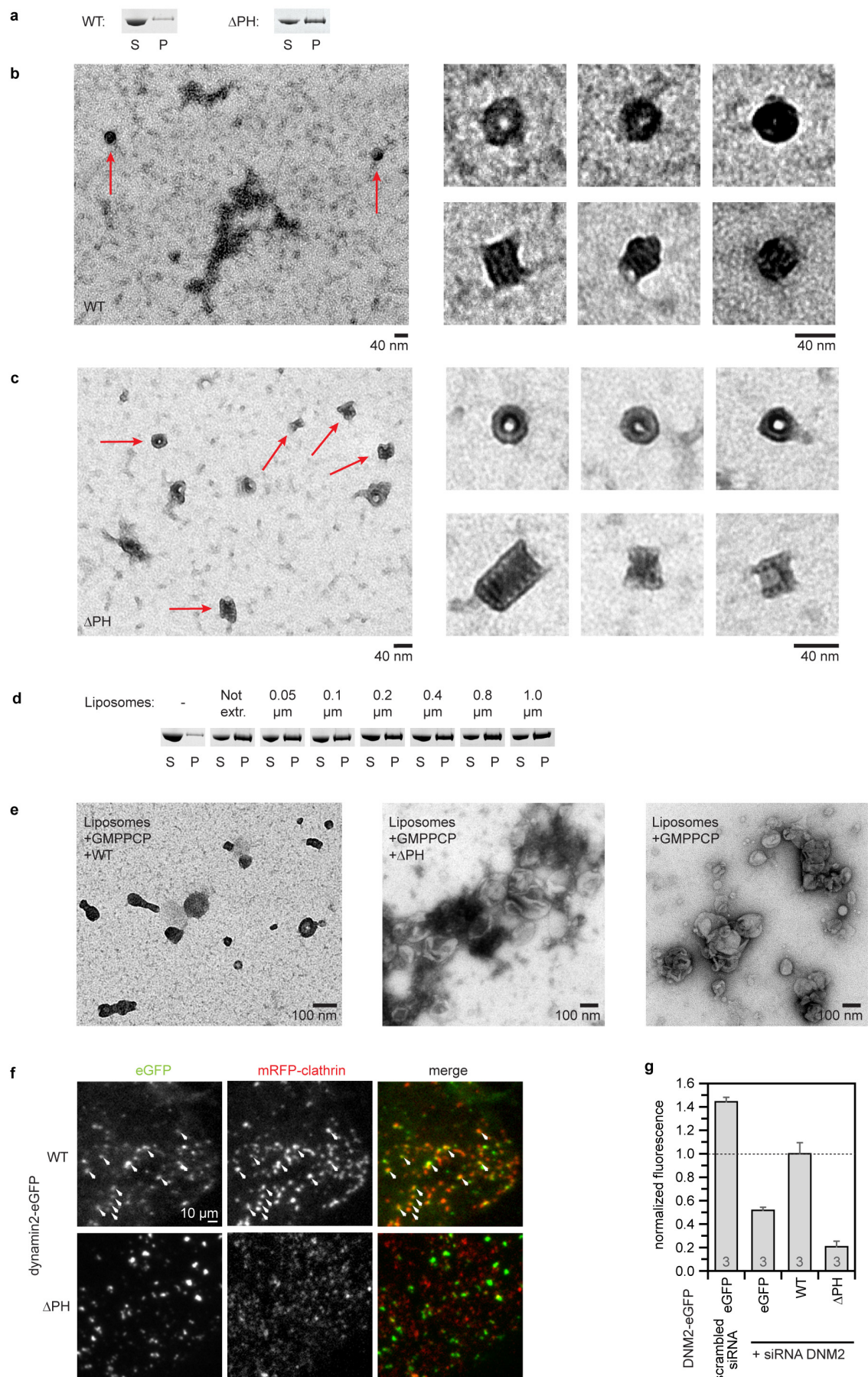


**Extended Data Figure 4 | Localization of the PH domain in the tetramer.**

**a**, Superposition of an outer dynamin molecule (magenta) and an inner molecule (green) of the dynamin 3 tetramer. The comparison reveals an  $\sim 40^\circ$  rotation of the G domains and BSEs. Furthermore, the PH domain is visible only in the outer molecule. **b**, Superposition of the stalk and PH domain in dynamin 1 (grey) and dynamin 3 (magenta). **c**, Connectivity of PH domain and stalk in the outer molecule. Shown are the stalk and PH domain of an outer molecule (magenta) and the stalk of the corresponding inner molecule (green) from a dimer. Since the gap of  $\sim 58$  Å between V629 of the PH domain and P643 of the inner stalk is too large to be spanned by the missing 13 residues (grey dashed line), we can unambiguously assign the PH domains in dynamin 3 to the outer stalks (black dashed lines). All other potential connections including molecules from the second dimer or symmetry-related tetramers

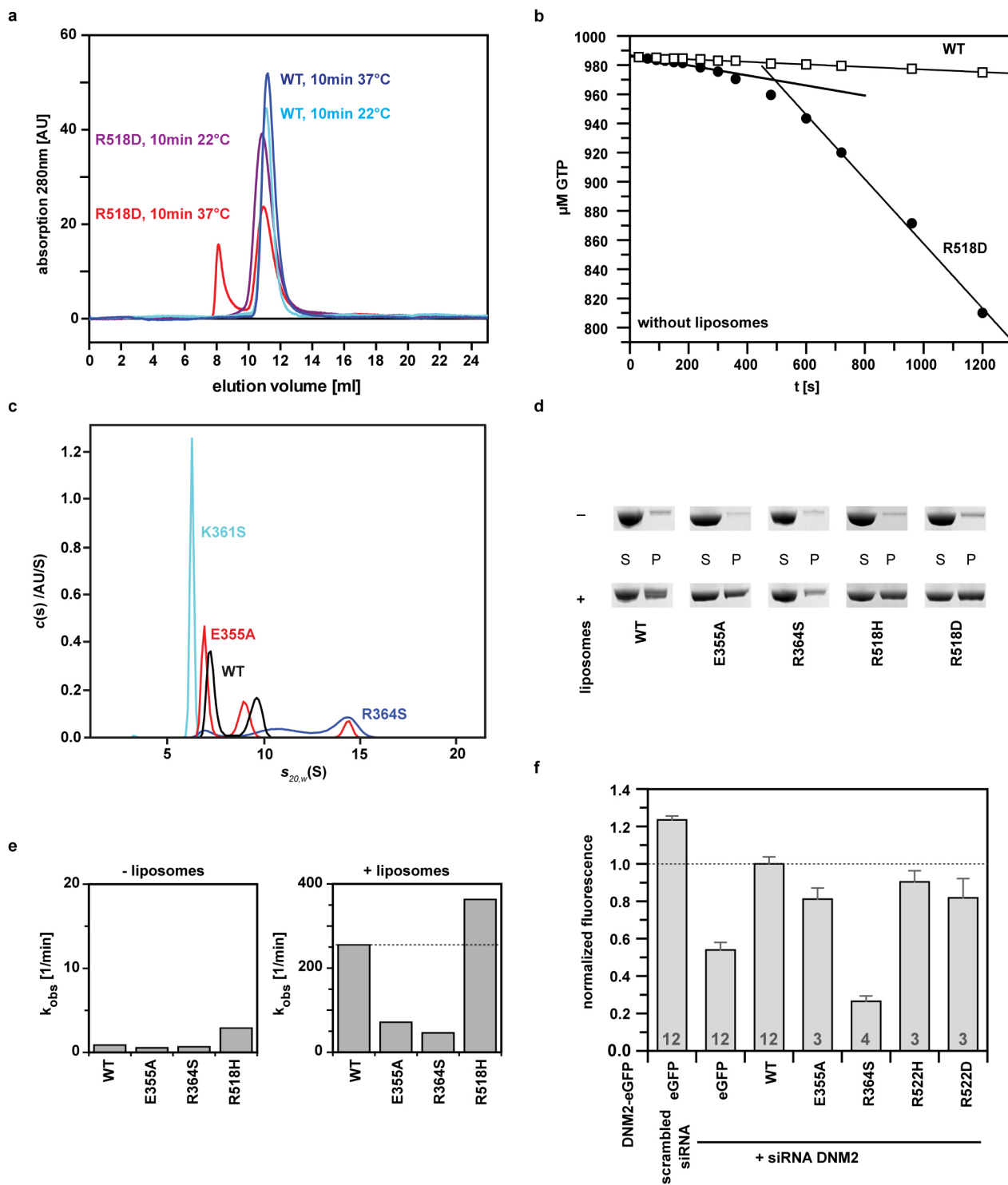
span even larger distances (not shown). In the crystal structures of dynamin 1, an unequivocal assignment of the PH domain to a specific stalk was not possible, due to the long unresolved linker regions between the stalk and the PH domains. Concomitantly, the impact of the interface between stalk and PH domain has not been generally recognized<sup>53</sup>. **d**, The outer PH domains are clearly defined in the electron density (left panel), whereas no density for a PH domain is observed in the equivalent position at the inner stalks (right panel). The density visible in the right panel corresponds mainly to a G domain from a symmetry-related molecule. The  $2F_o - F_c$  electron density is contoured at  $1.0\sigma$ . **e**, Modelling of a PH domain (grey) relative to an inner stalk (green) in the same geometry as seen in the outer molecules leads to steric clashes (black oval) with an adjacent stalk (blue).





**Extended Data Figure 5 | The PH domains regulate oligomerization of dynamin.** **a**, In the absence of liposomes, a dynamin 3 variant lacking the PH domain ( $\Delta$ PH) was sedimented more efficiently than wild-type dynamin 3 (WT). Both  $\Delta$ PH and wild type lacked the PRD. The proteins were sedimented by ultracentrifugation after 20 h of incubation at low salt concentrations (60 mM NaCl) in the presence of the non-hydrolysable GTP analogue GMPPCP. S, supernatant; P, pellet fraction. **b**, **c**, Representative negative-stain electron micrographs of wild type (**b**) and  $\Delta$ PH (**c**) under the same conditions as in **a**. For each protein, at least eight micrographs were recorded. Both constructs showed oligomeric ring structures, similar to structures seen for full-length dynamin<sup>7</sup>. Our data indicate that oligomerization of dynamin does not require membrane binding, but membrane binding requires oligomerization (Fig. 2). **d**, In liposome co-sedimentation assays, dynamin 3

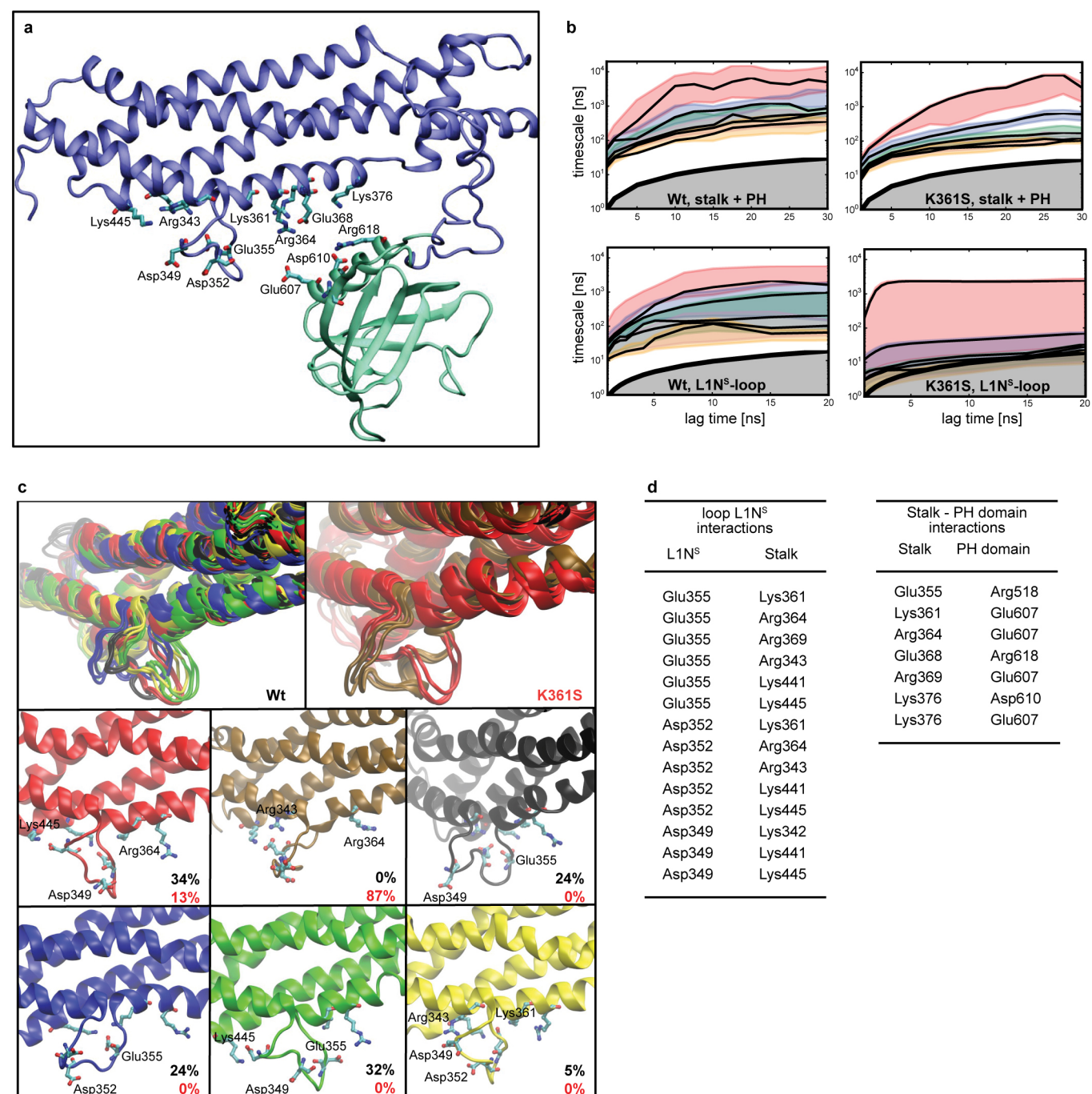
bound to Folch liposomes independently of their size. Not extr., not extruded. **e**, At physiological salt concentrations (150 mM NaCl), dynamin 3 efficiently tubulated unfiltered Folch liposomes. In contrast,  $\Delta$ PH did not decorate the liposome surface and did not induce liposome tubulation. For each setup at least 12 micrographs were recorded. **f**, When expressed in HeLa cells, dynamin 2( $\Delta$ PH) formed large cytosolic aggregates that did not co-localize with mRFP-clathrin. Arrowheads indicate co-localization for wild-type dynamin 2. Shown are magnified insets of representative images from at least 20 individual cells, acquired by TIRF microscopy. **g**, Dynamin 2( $\Delta$ PH) was dominant-negative in transferrin uptake assays. Data shown represent mean  $\pm$  s.e.m., the number of independent experiments is indicated in the bar. Raw data for **a** and **d** is available in Supplementary Information.



**Extended Data Figure 6 | Mutational analysis of the interface between PH domain and stalk.** **a**, Analytical gel filtration analysis for wild-type dynamin 3 and the mutant R518D. The proteins were pre-incubated for 10 min at 22 °C or 37 °C. When pre-incubated at 37 °C, only R518D showed a higher molecular weight species. AU, arbitrary units. **b**, Intrinsic GTPase activity of wild-type dynamin 3 and the mutant R518D at 37 °C in the absence of liposomes. The lines represent linear fits of GTP hydrolysis versus time. For R518D, a biphasic behaviour of the GTPase activity was apparent (for wild type:  $k_{\text{obs}} = 0.5 \text{ min}^{-1}$ ; for R518D:  $k_{\text{obs1}} = 2.2 \text{ min}^{-1}$  and  $k_{\text{obs2}} = 13.3 \text{ min}^{-1}$ ). This biochemical behaviour is reminiscent of dynamin 1 mutants in the PH-domain–stalk interface that show increased oligomerization and GTPase rates when incubated at 37 °C<sup>20</sup>. Perturbations in this interface appear to promote oligomerization of dynamin, pointing to an autoinhibitory function of this interface for oligomerization. The average of two independent measurements is shown with deviations ranging from 0% to 0.05% for wild type and 0% to 0.62% for R518D. **c**, Analytical ultracentrifugation experiments for the indicated dynamin 3 variants, as in Fig. 2a. For the mutant K361S that sediments as a single species, a molecular mass of 164 kDa could be obtained from  $c(s)$  analysis,

indicating that this mutant forms dimers in solution. **d**, Liposome co-sedimentation analysis for the indicated mutants. S, supernatant; P, pellet fraction. **e**, GTPase activity of the indicated mutants in the absence and presence of liposomes. Shown is the average of two independent measurements, with deviations ranging from 1% to 11%. **f**, Ability of dynamin 2 mutants to rescue defective CME of transferrin in absence of endogenous dynamin 2. The assay was performed as described in Fig. 2d. R518 in dynamin 3 corresponds to R522 in dynamin 2 and the R522H mutation in dynamin 2 is implicated in centronuclear myopathy. Data shown represent mean  $\pm$  s.e.m., the number of independent experiments is indicated in the bar. Note, we generally observed that the GTPase experiments were the most sensitive indicators of structural perturbations induced by mutations. Compared to membrane binding assays, GTPase assays appear to be more sensitive to the actual architecture of the dynamin oligomer and alterations induced by point mutations. Transferrin uptake assays could be influenced by cellular factors, such as BAR-domain protein that may stabilize mutant dynamin forms with deficits in oligomerization. Raw data for **d** is available in Supplementary Information.

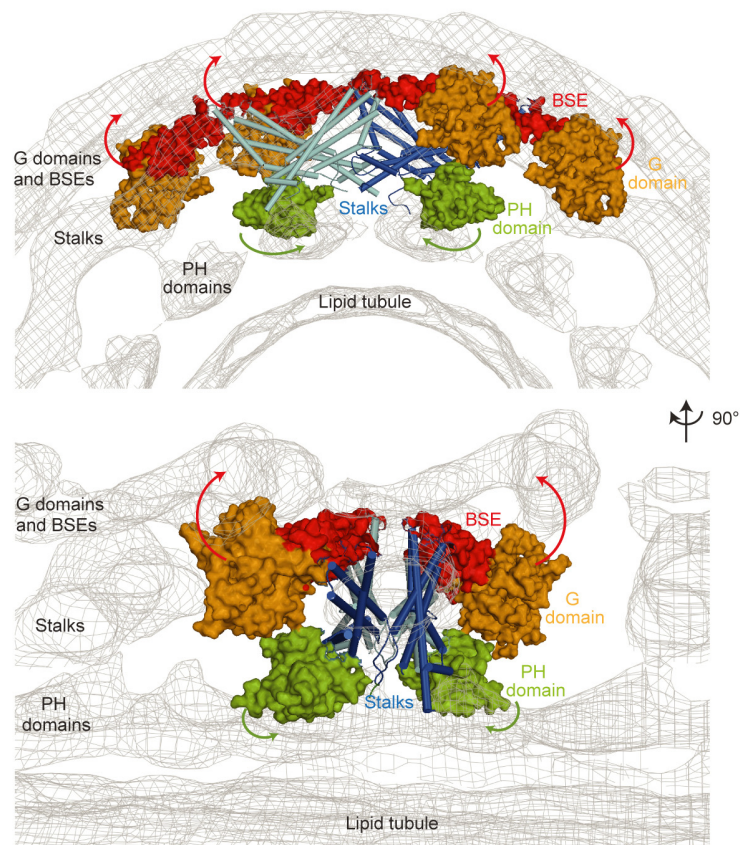




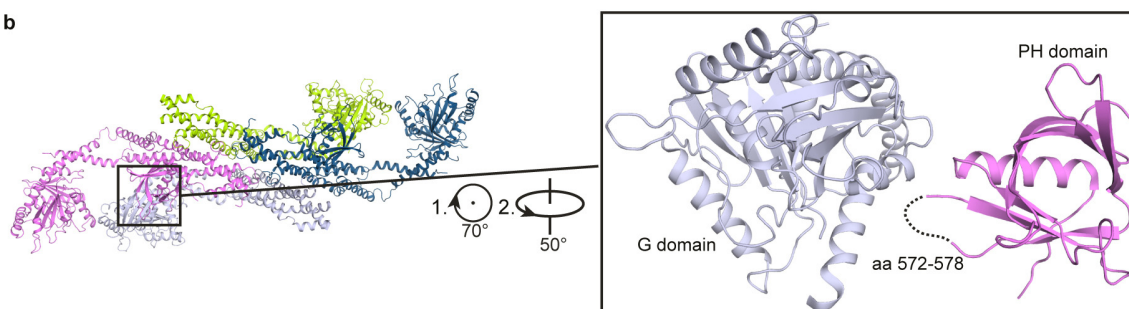
**Extended Data Figure 7 | Molecular dynamics simulations and Markov models.** **a**, The PH-domain-stalk interaction is characterized by a number of mainly polar interactions. The represented conformation is one of the starting structures (setup 2) for the MD simulations and quickly converts into one of the metastable conformations shown in Fig. 3. **b**, Relaxation timescales of different constructs as a function of lag time computed from Markov models. The timescales of all models (black) have converged at a lag time of about 20 ns within statistical uncertainty (colour-shaded regions),

indicating approximate Markovianity. The grey area indicates the region with lag times larger than relaxation timescales. **c**, Top: intrinsic conformation dynamics of the L1N<sup>s</sup> loop shown for the wild type (black) and the mutant K361S (red). Bottom: six metastable conformations and their equilibrium probabilities of the L1N<sup>s</sup> loop (setup 3) for the wild type (black) and mutant K361S (red) computed from the Markov model. **d**, Residue pairs used to characterize the L1N<sup>s</sup> loop and stalk-PH domain interactions.

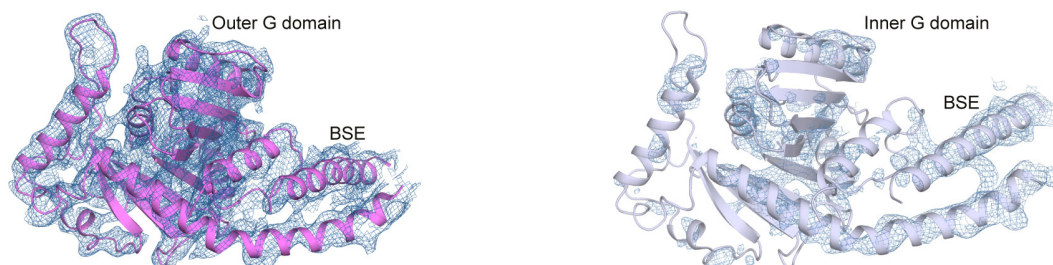
**a**



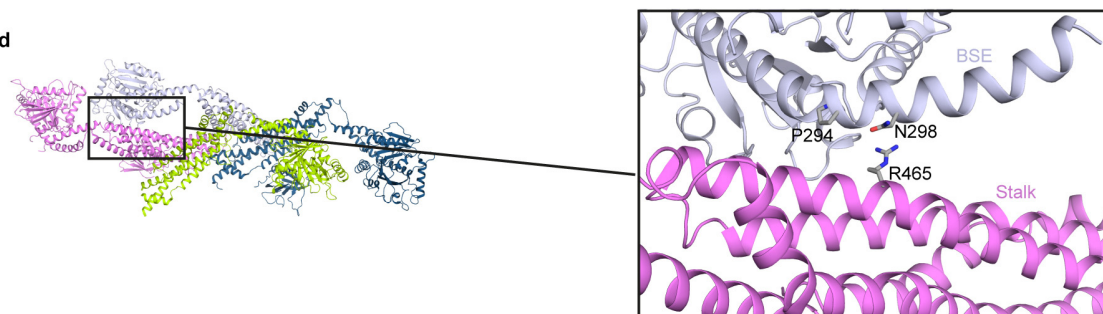
**b**



**c**

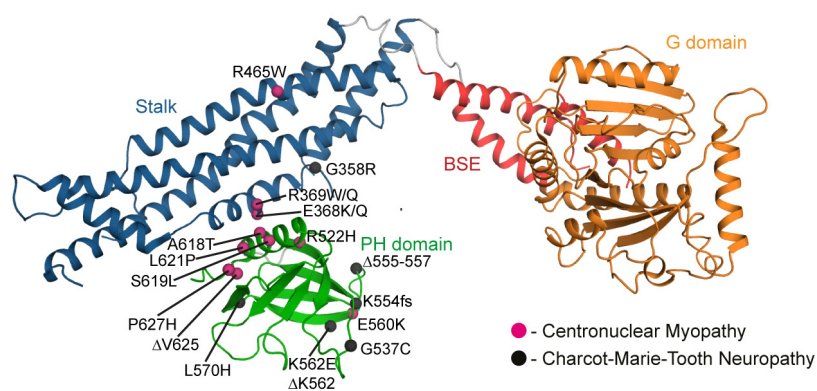


**d**



**Extended Data Figure 8 | Interactions of the G domain, stalk and BSE in the tetramer.** **a**, Two views on a fitting of the dynamin 3 tetramer crystal structure into the EM density of non-constricted oligomerized dynamin 1 (ref 24). The positions of the inner G domains are shown in all four molecules since the outer G domains in our crystals are stabilized by crystal contacts. Apparently, membrane binding and oligomerization is associated with major movements of the G domain, BSE and the PH domain (indicated by arrows). **b**, A loop of the outer PH domain and an inner G domain are in close proximity. **c**, The outer G domains (left), but not the inner G domains

(right), are well defined in the electron density. The  $2F_o - F_c$  electron density is contoured at  $1.0\sigma$ . The weak electron density for the inner G domains and the resulting uncertainty in determining the contact sites prevented us from analysing this interaction in more detail. **d**, The BSE of an inner monomer (grey) interacts with the stalk of an outer monomer (magenta). This contact involves R465 which is mutated to tryptophan in some centronuclear myopathy patients. The R465W mutation leads to hyperactive dynamin<sup>4</sup> that fragments the T tubule network in mouse-myoblast-derived myotubes and *Drosophila* body wall muscle (Y.-W. Liu, personal communication).



**Extended Data Figure 9 | Disease-relevant mutations in dynamin.** Localizations of mutations leading to Charcot-Marie-Tooth neuropathy (black balls) and centronuclear myopathy (pink balls) are plotted onto a dynamin 3 monomer. Colour code as in Fig. 1.



Extended Data Table 1 | Data collection and refinement statistics

**Data collection**

Space group	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub> , 1 tetramer / ASU
Cell dimensions	
<i>a</i> , <i>b</i> , <i>c</i> (Å)	97.70, 98.00, 401.52
Resolution (Å)*	3.70 (3.70-3.80)
<i>R</i> <sub>sym</sub> (%)*	7.0 (130)
<1/ $\sigma$ ( <i>I</i> )> *	16.6 (1.8)
Completeness (%)*	99.6 (99.3)
Redundancy	7.3

**Refinement**

Resolution (Å)	49.47 – 3.7
No. reflections	42,058
<i>R</i> <sub>work</sub> / <i>R</i> <sub>free</sub> (%)	23.2 / 27.8
No. of protein atoms	18,654
averaged B-factor protein (Å <sup>2</sup> )	212
R.m.s deviations	
bond lengths (Å)	0.004
bond angles (°)	0.889

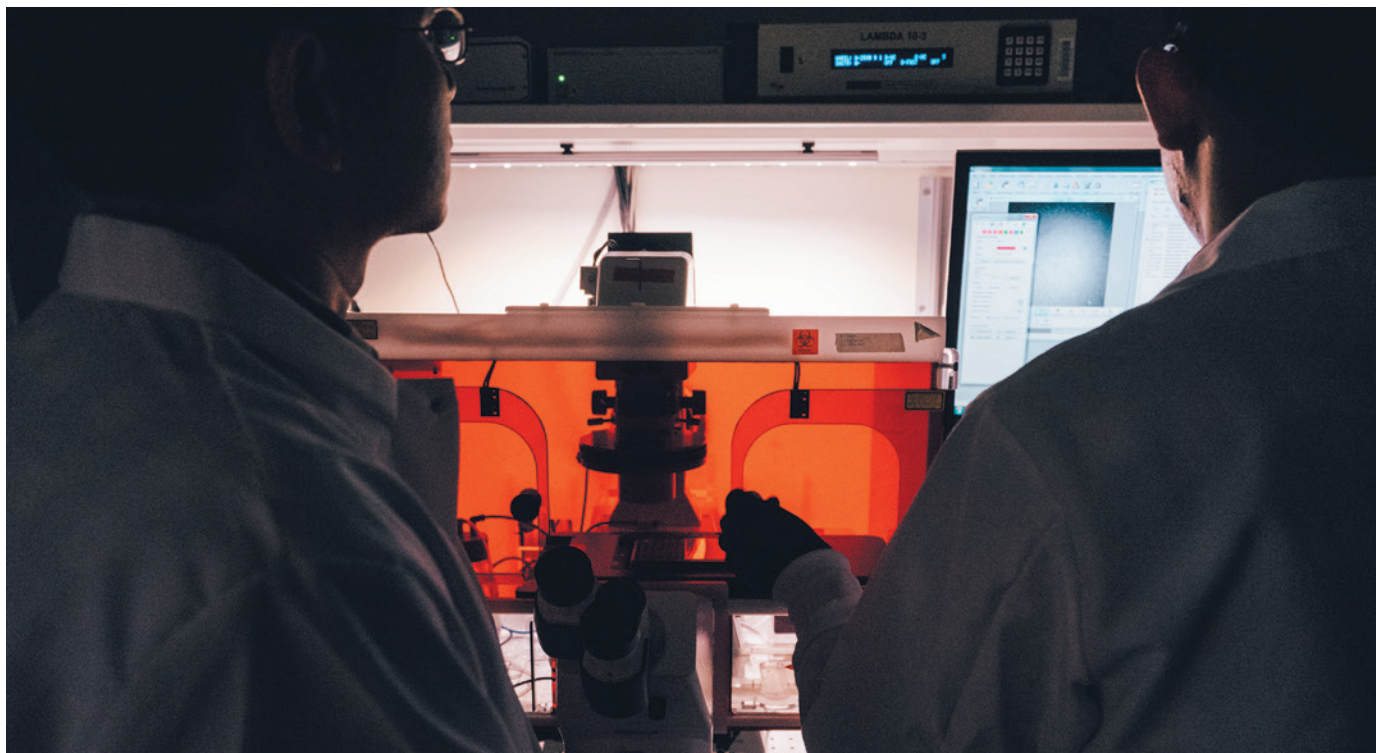
\* Data in highest resolution shell are indicated in parenthesis.

## TECHNOLOGY FEATURE

# THE CELL MENAGERIE: HUMAN IMMUNE PROFILING

*Cutting-edge tools and analyses are digging deeper than ever before to unveil the intricacies of the diverse human immune system.*

LILLIE PAQUETTE, MIT SCHOOL OF ENGINEERING



Advanced technologies enable researchers at the Massachusetts Institute of Technology to observe individual immune cells attacking tumour cells.

BY MARISSA FESSENDEN

Vaccines save lives — but they don't always work. Take the annual influenza shot: by some estimates, flu vaccines are only 50–70% effective even when well matched to the virus strains in broad circulation. Despite all the research, scientists still cannot predict whether a given vaccine will work for any specific person.

Learning to make vaccines that protect more people means getting a better handle on the immune system — a bewildering militia of cells that communicate to detect and destroy pathogens. So far, attempts to parse the system's complexity have involved work on mice, rats, rabbits, dogs, non-human primates and even lampreys and sea urchins. Yet results do not always translate to the one species that

medicine cares most about. “There has been a vast zoo of animal models, but the one animal model we haven't yet exploited is us — *Homo sapiens*,” says Bali Pulendran, an immunologist at Emory University in Atlanta, Georgia.

Now, researchers are tackling the most difficult animal to study as never before. Advances in technology are helping scientists to dive deeper into the inner workings of single cells and carry out analysis on greater numbers of cells at once. Efforts in data analysis, sharing and collaboration promise to enable work that is too expensive for individual labs. Ultimately, researchers hope to bring fresh insights to the clinic to protect and treat people using the power of an individual's own immune defences.

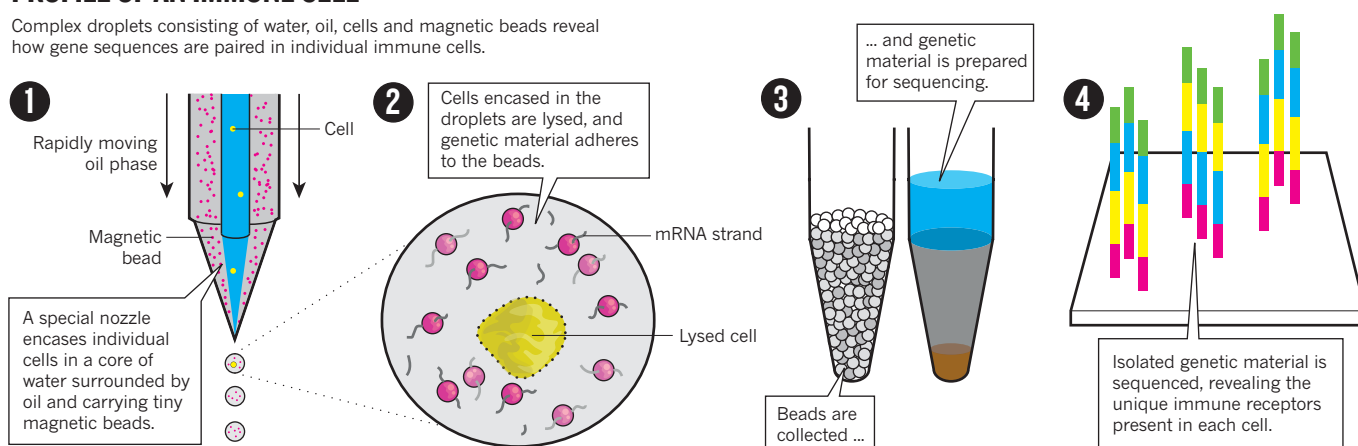
The human immune system is incredibly diverse. Each class of immune cell is actually an army of subtypes. The elite forces — the

lymphocytes, which recognize specific pathogens or wayward body cells — consist of natural killer (NK) cells, which quickly dispatch infected or cancerous cells, and B and T cells, which bear receptors on their surfaces designed to recognize specific invaders. But B and T cells break down further: there are regulatory T cells, T helper cells, memory B cells, naive B cells and more, each with its own unique role. These lymphocytes coordinate in turn with cells such as macrophages and monocytes, which are further specialized for other functions.

Diversity manifests between people, too. Even identical twins vary in terms of the exact molecules and cell profiles that fight off disease. From an evolutionary point of view, variability ensures that some members of a species will survive a deadly disease outbreak — but it confounds researchers. ►

## PROFILE OF AN IMMUNE CELL

Complex droplets consisting of water, oil, cells and magnetic beads reveal how gene sequences are paired in individual immune cells.



► Gender, ethnicity, genetic background and disease history all affect a person's immune response in unpredictable ways. They influence whether a vaccine will work, and whether someone has allergies or an autoimmune disease — both resulting from an overactive immune system — or whether a person will develop cancer, which is caused in part by an inattentive system that fails to remove errant cells.

### VACCINES UNVEILED

Instead of seeing confusion in such diversity, researchers such as Pulendran see opportunity. With the right combination of sophisticated technologies and data analysis, human variation can offer a natural experiment in what underlies an effective immune response.

This reasoning led Pulendran and his team to some groundbreaking research on why a vaccine for yellow fever works so well. Since immunization against the sometimes-deadly tropical disease began in 1937, only 12 cases have been reported among the hundreds of millions of people immunized.

Scientists have long known that the vaccine spurs the body to produce T cells that can kill cells infected with the yellow-fever virus — but they did not know how. In 2009, Pulendran's team published an analysis<sup>1</sup> of changes in the state, number and types of immune cells in the body before and after vaccination. The group found that quantities of a protein called EIF2AK4 spike in key immune cells (mainly dendritic cells, which help T cells to identify invaders) just days after vaccination. The higher the spike in protein levels, the more anti-yellow-fever T cells are later produced.

The close correlation suggests the existence of components that foster strong immune responses — at least for the yellow-fever vaccine. Pulendran and his colleagues<sup>2</sup> have since discovered other proteins that predict similarly strong responses to vaccines for flu and meningococcal disease. Now, they are linking these types of marker to subpopulations of cells and classifying variation across individuals.

One major reason that immune responses vary is the vast collections of receptors on the

surfaces of T and B cells, which correspond to antibodies that are secreted by the latter cells. To produce a near-endless assortment of these Y-shaped molecules, lymphocytes shuffle their genes as they mature. The myriad receptors and antibodies that result enable the immune system to recognize many different pathogens.

Researchers want to sequence genes for these receptors to work out what makes a potent immune response, and so gain clues for developing vaccines and for designing therapies that could spur the immune system to fight cancer.

But because each receptor is made from proteins encoded by at least two types of separately shuffled gene segment, sequences alone are not enough. Researchers must also learn how these proteins are paired in an individual cell — and which combinations show the most promise for fighting disease.

At the University of Texas at Austin, chemical engineer George Georgiou has tackled this challenge by studying B cells one at a time. He and his team<sup>3</sup> first encase individual cells in complex droplets: ones with an aqueous core that preserves the cell's genetic material; an outer oil layer to keep the cells separated; and magnetic beads that allow researchers to manipulate each droplet and so capture and extract the genetic material from individual cells. These data can reveal the antibody repertoire elicited by various stimuli — crucial information for designing vaccines (see 'Profile of an immune cell').

Georgiou's group hopes to publish manual-like methods so that others can use the technique. Investigators who are unfamiliar with it should prepare themselves for a steep learning curve, he says: "Every method, especially new methods from academic labs, has some nuances." But for those who are willing to put in the time and elbow grease, precise answers about individual immune cells await.

Another sequencing approach relies on

specially formulated beads to tag individual cells before DNA analysis. The tags fuse with the cells' genetic material and function as barcodes that can be traced back to the original cell, even when cells are analysed in pools. Hedda Wardemann, an immunologist at the Max Planck Institute for Infection Biology in Berlin, has used this strategy to analyse genes encoding paired receptor proteins in more than 46,000 B cells at a time<sup>4</sup>.

Most microfluidics sequencing platforms are developed by individual labs, such as Georgiou's or Wardemann's, that have engineering know-how. But as the field grows, companies are getting into the game. One of the biggest players in the microfluidics field is Fluidigm in South San Francisco, California.

This September, the company began to ship high-throughput chips, which can be used on the company's C1 microfluidics platform to interrogate genomes of 800 individual cells in a single 6.5-hour run. Although it has a much lower throughput than Georgiou's droplet method (which can process 6 million B cells in a day), Fluidigm's technology requires less expertise. The company plans to increase throughput to nearly 100,000 cells per run in the near future.

### PERTURBED POPULATIONS

In addition to profiling individual cells (see 'Nanoarenas for cell attacks'), researchers want to track how cell populations change in response to vaccination or infection. To identify specific cell types, the scientists rely on protein markers studded on the cells' surfaces.

For example, two markers dubbed CD4 and CD8 both show up on certain types of memory T cells — but CD8 is also on NK cells, and CD4 is on monocytes and dendritic cells. So, to measure only memory T cells, researchers may need to screen for three different markers. To isolate an even more-specific subset, the number of markers must increase.

Conventionally, researchers have relied on a cell-profiling technology called flow cytometry, in which coloured, fluorescent proteins are attached to specific cell markers so that combinations can be easily detected

SOURCE: REF. 3



and the cells scored or sorted. But overlaps in colour spectra generally limit analyses to as few as a dozen markers.

The latest iteration of cell-profiling technology — mass cytometry — uses rare-earth metals instead of fluorescence and can detect more than 40 markers. Because mass cytometry can identify so many cell types in a single sample, more types of experiments can be done.

Studies in babies, for example, are key to understanding the immune system's development. But infants generally cannot tolerate blood withdrawals of more than 4–5 millilitres — and even simple flow-cytometry experiments can require more than 10 ml. Mass cytometry, by contrast, can run on less than 4 ml.

Mark Davis, a molecular immunologist at Stanford University in California, used mass cytometry to track hundreds of parameters — including 72 different immune-cell populations — in the blood of 210 twins. His team found<sup>5</sup> that much of the variation between people's immune systems can be attributed to environmental factors, rather than to genetic ones. Without mass cytometry, this work would have been too complex to perform, he says.

DVS Sciences, now a part of Fluidigm, has invented a mass cytometer called the CyTOF for use in cell profiling. The latest version (as well as upgrades for the older system) boosts sensitivity and sample-processing speed, and can run multiple samples at a time.

But these technologies are expensive. The June version of the CyTOF — the 'Helios' system — starts at roughly US\$500,000, not counting service contracts. At Stanford, Davis and other researchers rely on shared facilities.

## ASSEMBLING THE PIECES

Although scientists are making progress, many tools have been slow to reach the clinic, says Padmanee Sharma, a physician-scientist at the University of Texas MD Anderson Cancer Center in Houston. Every new clinical technology needs standards and quality assurances, which require extensive testing to establish. Clinical trials are only now adopting procedures that might help clinicians to track their patients' immune responses and feed in to treatment decisions.

Communication is another bottleneck. Information is accumulating rapidly and needs to be shared by collaborators as diverse as statisticians, clinicians, basic biologists and technologists. Coordinating research that involves human participants places huge demands on logistics, resources and expertise, and one major effort to facilitate such work is the Human Immunology Project Consortium (HIPC) funded by the US National Institutes of Health (NIH). The HIPC doles out grants to advance methods, and endeavours to extend the fruits of researchers' labour to all.

The consortium offers an online data-analysis and management platform called ImmuneSpace, which helps researchers to place data

# Nanoarenas for cell attacks



In addition to tracking populations of immune cells, researchers want to know how they interact.

Christopher Love, an immuno-engineer at the Massachusetts Institute of Technology in Cambridge, is using microfluidics to probe how individual immune cells cooperate with each other. His lab engineers devices that he describes as “essentially ice-cube trays”: each

well in the tray holds sub-nanolitre volumes, as opposed to the tens of microlitres held by wells in more-conventional plates.

Using these tiny arenas to watch natural killer (NK) cells home in on leukaemia cells, the team has discovered<sup>7</sup> — unexpectedly — that even a single NK cell will attack a cell that does not belong. In the past, researchers suspected that NK cells coordinated their actions through secreted chemical signals, but now it seems that such cooperation may be necessary only among larger cell groups.

Love and his team hope to map their understanding of interactions at this single-cell level to the immune system as a whole, and potentially compare healthy individuals with those who have cancer. “With these technologies, first you ask: can we define normal?” he says. “Then you can think about heterogeneity in disease.” **M.F.**

in a long-term archive called the Immunology Database and Analysis Portal (ImmPORT), also funded by the NIH. The HIPC is spearheading efforts to standardize procedures for commonly performed assays in cytometry as well as alternate methods of immune profiling, such as measuring antibodies in serum samples.

Another emerging need is for techniques for easy cross-analysis of many data types, says Steve Kleinsteinst, a computational immunologist at Yale University in New Haven, Connecticut. “There’s a lot of subtlety in the data, and it’s very easy to pick up a piece of code or tool that somebody put out there on the web, run it with your data and get a plot that looks interesting — but that’s a very dangerous thing to do,” he says.

To help solve this problem, Kleinsteinst and his group have developed software called the Repertoire Sequencing Toolkit (pRESTO)<sup>6</sup>, which offers a way to process, annotate and correct raw sequencing data from high-throughput platforms such as Illumina. It also allows researchers to run their data in different computing environments and then return to the pRESTO environment.

A separate tool, a web portal known as the VDJServer, is in beta-testing after launching in April. It offers the ability to analyse B- and T-cell-receptor data, with the goal of providing an intuitive interface for users who have not done any programming, says project leader Lindsay Cowell, a bioinformatician and immunologist at the University of Texas Southwestern Medical Center in Dallas. The server will incorporate more analysis tools into the portal as they become available (Kleinsteinst’s pRESTO is already embedded). Moreover, the portal lets researchers share data and even tap into the computing power of the

Texas Advanced Computing Center at the University of Texas at Austin.

There is still an acute need for human immunology-specific data repositories, notably for T- and B-cell-receptor sequencing data, says Jamie Scott, a molecular immunologist at Simon Fraser University in Burnaby, Canada, who is co-leading an effort to share such data.

But perhaps the biggest block is a basic one: a dearth of training. Most analysis requires some programming skills, says John Tsang, head of computational systems biology for the Trans-NIH Center for Human Immunology in Bethesda, Maryland. For now, most tools are limited to the specialist, he says; collaboration with those who can understand the programming is still the best way forward.

Creating more collaborations should, in turn, help to ensure that the tools truly further basic knowledge and translate into practical applications. “It is very attractive to apply the latest gee-whiz ‘omics’ technology to measure things,” says Pulendran. “But I think we need to go beyond measuring and accumulation of data — to knowledge and to understanding.” ■

**Marissa Fessenden** is a science journalist and illustrator based in Bozeman, Montana.

1. Querec, T. D. *et al. Nature Immunol.* **10**, 116–125 (2009).
2. Li, S. *et al. Nature Immunol.* **15**, 195–204 (2014).
3. DeKosky, B. J. *et al. Nature Med.* **21**, 86–91 (2015).
4. Busse, C. E., Czogiel, I., Braun, P., Arndt, P. F. & Wardemann, H. *Eur. J. Immunol.* **44**, 597–603 (2014).
5. Brodin, P. *et al. Cell* **160**, 37–47 (2015).
6. Vander Heiden, J. A. *et al. Bioinformatics* **30**, 1930–1932 (2014).
7. Yamanaka, Y. J. *et al. Integr. Biol.* **4**, 1175–1184 (2012).



# CAREERS

**GENE EDITING** One investigator's experience launching a lab with CRISPR **p.415**

**NETWORKING** Tips for introverts engaging in a professional setting [go.nature.com/9oqhff](http://go.nature.com/9oqhff)

**NATUREJOBS** For the latest career listings and advice [www.naturejobs.com](http://www.naturejobs.com)



## WORK ENVIRONMENT

# When labs go bad

*A toxic relationship between junior scientist and adviser can quickly turn career prospects sour.*

BY CHRIS WOOLSTON

There is no crying in baseball, according to a famous quote from the 1992 film *A League of Their Own*. But there is most certainly crying in science, says Isaiah Hankel, a former cell biologist turned author and career coach. He admits shedding a couple of tears in a bathroom cubicle after his graduate adviser screamed at him in front of the entire lab — all while another principal investigator (PI) looked on. “It was the craziest thing,” he says.

But it did not come entirely out of the blue. During his fifth year of study, Hankel had been promised an industry job — under the condition that he get his PhD first. Unfortunately, his PI was not on board with the plan. “He totally withdrew his support,” he says. “I wanted to map out exactly what I needed to do for graduation, but he would never nail it down.”

Like Hankel, many junior researchers come to realize that their relationship with their PI — the one person who is most in control of their careers — is not working out. “I’ve seen a lot of

situations where people are having problems with their supervisors,” says Sarah Blackford, head of education and public affairs for the Society of Experimental Biology, headquartered in London. “People get very emotional, and things can escalate.” Blackford, who is based at Lancaster University, UK, and advises junior researchers throughout Europe, says that postdocs and graduate students in broken labs must work out the crucial next step. Are they going to endure a bad situation? Are they going to find a way to mend the relationship? Or are they going to jump ship?

Whatever the decision — endure, repair or escape — the conflict will probably become a career turning point. Junior researchers who run afoul of their PIs may feel stuck, and they could end up with one fewer letter of recommendation than they had originally counted on, but that does not mean that their science days are over. With a positive attitude, a knowledge of institutional policies and some objective, well-placed allies, it is possible to move on — to academia or beyond.

## A CHANGE IN TACK

Hankel quickly realized that long hours and dedication were not going to be enough to break the impasse with his PI. “Working extra hard is exactly what my PI wanted,” he says. “I was getting more data for him. But if they aren’t going to give you a target to hit, you can’t keep spinning your wheels.” Instead of working harder, Hankel used some of his paid time off, giving himself time to make a plan. He started attending conferences, which he paid for out of pocket. That sort of networking, he says, can be especially important in times of conflict. He kept daily records of his interactions with his PI, and he saved all of the relevant e-mails.

Most importantly, he set up meetings with his department head and several deans, and discussed with them his need for a clear path to graduation. He also consulted the school’s official graduate-school manual, which gave him a major source of leverage. Among other pronouncements noted in the manual, students were expected to graduate within five years, and advisers were supposed to actively support their students’ progress. Prompted by the meetings, his adviser finally told him the exact steps that he needed to take to finish his dissertation. With an exit plan in place, Hankel was able to get his degree about a year after all of the trouble started.

Conflicts with senior scientists can be especially bewildering for PhD students, says ►

## THE BEST DEFENCE

*Look before you leap*

Many junior researchers who find themselves at odds with their advisers could have avoided trouble with a little preliminary research. For PhD students, it is helpful to find someone who has a history of turning trainees into scientists, says career adviser Karen Kelsky in Eugene, Oregon.

“Be a good detective. Check with other graduate students and postdocs, and look at the track record,” she says. The statistics will tell the story — many ‘bad’ advisers have never guided a doctoral student through the point at which he or she actually earned a degree. Of course, some principal investigators are too new to have much history. In those cases, Kelsky says, students should check with prospective advisers to make sure that they are committed to helping students to earn their degrees.

Postdocs too often take a scattershot approach to finding a lab, says Sofie Kleppner, an assistant dean in the office of

postdoctoral affairs at Stanford University in California. “Some of them will spam the entire university looking for a position,” she says. “They spam me, and it’s been a long time since I’ve had a lab.”

Instead, they should conduct a much more focused search for a lab that is compatible with their personality, rather than just their scientific interests. She recommends that postdocs give a talk to the principal investigator and members of a prospective lab, creating an important opportunity for both sides to look for a good fit. In addition, they should set up an in-person chat with the adviser — and have lunch or dinner with other people in the lab. This is the chance to ask a question that could prevent a lot of future trouble: what is the worst thing about working in this lab? If the complaints run far beyond the normal scientific grumblings, it is better to keep looking. **C.W.**

► Karen Kelsky, a science job coach in Eugene, Oregon, and author of *The Professor is In: The Essential Guide to Turning your PhD into a Job* (Three Rivers Press, 2015). PhD students do not always have the interpersonal experience to handle rocky relationships, she says, and they are often unprepared for the rigid hierarchy of academia (see ‘Look before you leap’). “There are some aggressive advisers who like the power and just want to see a person get destroyed,” she says. Instead of letting a student defend a thesis and receive a PhD, they ask for one more rewrite or one more experiment, not because the work is crucial, but to remind the student who is really in command, she says. “That’s probably the most common story I hear,” she says.

In many cases, students can get their freedom by putting their head down and meeting every request, even if it seems wrong or unhelpful. “That’s what ended up happening to me,” says Kelsky, who has a PhD in cultural anthropology. “I revised my dissertation by taking out everything my adviser hated and putting in everything she liked.” As they approach the finish line, she says, students should think less about their literary legacy and more about making their PI happy. “A lot of graduate students are obsessed with their dissertations, but the fact is that nobody is going to read them. They shouldn’t get so worked up.”

## SEARCH FOR ALLIES

One cognitive scientist, who asked not to be named, received her PhD from a prestigious university on the US West Coast. She says that her relationship with her PI fell apart in the fourth year of a five-year programme, a

particularly vulnerable time in her education. A combination of misdeeds, misunderstandings and hurt feelings left her wondering whether she should abandon the programme and start again. Among other questionable behaviours, her adviser seethed when she did some work with a rival lab during her adviser’s sabbatical. When her adviser gave one of her projects to another student, she felt the relationship was irretrievably damaged. But instead of quitting her PhD programme, she had coffee with a faculty member who helped her to look at the big picture. “She said

I shouldn’t throw away four years of work.” The same faculty member stepped up to become the co-chair of the student’s committee, a position from which she could ensure that the degree process would be fair and unbiased. “She made sure I wasn’t retaliated against,” says the cognitive scientist, who is now a tenure-track assistant professor at a US university.

For postdoctoral researchers, conflicts with PIs can cause a lot of soul-searching and career angst, says Sofie Kleppner, an assistant dean in the postdoctoral-affairs office at Stanford University in California. “It’s a huge issue if you’re in a lab and you feel like it’s the wrong lab for you,” she says. In her experience, postdocs often feel as if they and their advisers are not on the same page. “One of the biggest problems is mismatched expectations,” she says. “A postdoc might want to be independent, but a PI might be the type who likes to check in. That can cause a lot of frustration.”

In some cases, simple misunderstandings can cause a lot of tension. “A postdoc might tell me that they don’t want to go into academia, but they’re afraid to tell their PI,” Kleppner says. “And then the PI will say that he’s worried because the postdoc doesn’t seem cut out for academia.” The upside of simple misunderstandings, she says, is that they often have an equally simple solution: talking about it.

## THE ART OF CONVERSATION

As professional scientists, postdocs need to take a business-like approach to conflicts with their PIs, Blackford says. That means communication — and a lot of it. “You have to talk about the situation without getting personal,” she says. “Set up a meeting with a proper agenda.” Blackford adds that not all PIs are especially approachable or easy to talk to. If one-on-one conversations do not completely solve the problem, she recommends finding an



Science-career coach Karen Kelsky helps PhD students to navigate the job world.

KAREN KELSKY

impartial faculty member who is willing to offer confidential advice.

In some cases, Blackford says, discussing the situation with an objective ally can help disgruntled junior researchers to understand the true source of their discontent. “Some people can’t even put a finger on what’s gone wrong,” she says. “They just don’t feel respected, and then they have a crisis of confidence. It’s helpful to talk with someone who can tease out what you’re saying.”

Postdocs should develop on-campus allies who can serve as sounding boards and counselors. “I tell people to identify their peer support and mentors early,” Kleppner says. “You need someone who can advocate for you if something isn’t working out.” Adding another person to the conversation can be a quick way to find compromise and clarity, she says. “It’s basic ‘Conflict Resolution 101’.”

Not all conflicts can be resolved — some postdocs eventually decide to leave a lab for good. “These are high-powered people who don’t want to admit failure,” Kleppner says. “But it’s OK to admit it.” When it is time to leave, professionalism is more important than ever. She recommends explaining the decision to a PI in clear, dispassionate terms — the same tone that is needed when talking to other PIs about a possible job. Naturally, they will want to know why the last job did not work out, but they don’t want to be dragged into the drama. A postdoc who can clearly communicate why the last lab was not an ideal fit — without making any personal attacks on his or her former PI — will have a good chance of moving on. “You’re not going to ruin your reputation as long as you don’t ruin anyone else’s,” says Kleppner.

Hankel managed to leave academia with his reputation — and his degree — intact. As a career consultant, he now encourages other scientists to stand up for themselves even when the hierarchy is tipped against them. He notes that some scientists end up spending so many years doing their PhD and multiple postdocs that they barely have time to establish their careers before retirement. “Advisers hold the keys to people’s lives,” he says — which means that it is important to resolve disputes as quickly as possible and avoid spending too much time in a lab that will not promote a junior researcher’s progress. When a PI is not being supportive, Hankel says, early-career researchers have to prioritize their professional interests — even if that means hurt feelings, bruised egos and a change of venue. “It’s always appropriate to have self-respect,” he says. ■

**Chris Woolston** is a freelance writer in Billings, Montana.

A forced lab move can be a hassle. Find out how to handle it seamlessly in an upcoming issue of *Nature Careers*.

## TURNING POINT

# Martin Jinek

*Structural biologist Martin Jinek helped to launch the genome-modification craze that is upending biological research. Now running his own laboratory at the University of Zurich in Switzerland, Jinek describes how research is changing as CRISPR — a gene-editing tool with the potential to cheaply alter plants, animals and even human embryos — takes hold.*

### Did you set out to work on CRISPR after completing graduate school?

No. When I started as a postdoc in Jennifer Doudna’s group at the University of California, Berkeley, in 2007, we knew practically nothing about CRISPR, which stands for ‘clustered regularly interspersed palindromic repeats’. The first paper describing it as an adaptive immune system in bacteria came out early that year (R. Barrangou *et al. Science* **315**, 1709–1712; 2007). Although Doudna was one of the first to explore CRISPR, my original project was on the molecular mechanisms of microRNA. But the CRISPR field became more interesting, so I collaborated with some group members and finally began my own project working on Cas9, an enzyme that cuts DNA.

### When did it become clear that CRISPR was a game changer?

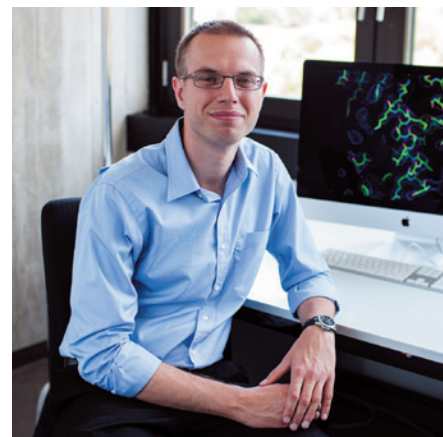
We were interested at first because it looked similar to RNA interference, in which RNA molecules inhibit the expression of genes. But the molecular machinery was intriguingly different. The wider implications — and its potential utility in genome research — came only after we learned that it cuts double-stranded DNA and is programmable, which made it even more interesting to work on.

### What is most surprising about this technology?

How quickly it has developed. Within six months of publishing a paper showing that CRISPR can be programmed (M. Jinek *et al. Science* **337**, 816–821; 2012), three labs — including ours — were using it as a genome-editing tool. Within 12 months, researchers were applying it to many cell types and organisms.

### How is CRISPR shaping your research agenda?

My goal is to understand how the system actually works. My resources are not unlimited, so I focus on what I do well — structural biology. Five of the ten people in my lab, which began in 2013, are aiming to gain a better structural understanding of the DNA-cutting mechanisms in CRISPR systems so that we can engineer the system to be more efficient and versatile. The CRISPR technology is finding



applications in basic-research labs, as well as in biotechnology and molecular-medicine labs, to potentially cure genetic disease or engineer organisms to make biofuels. I’m already using it to address other research questions.

### What did you take from your experience as a graduate student in a new lab?

I was the third PhD student in Elena Conti’s first laboratory, at the European Molecular Biology Laboratory in Heidelberg, Germany. She was a fantastic mentor, and being in her lab at an early phase of her career has shaped my own lab. She was a tough boss, but she taught me how to approach a scientific problem to find the right questions, and how to do good science to answer those questions.

### Has the public reaction to CRISPR had an impact on your work?

On some level, we anticipated it would be big. We just didn’t know how big. The wider societal and potential ethical issues associated with the use of CRISPR, especially those that relate to human-genome modification, have generated a lot of attention. The negative side of working in the CRISPR field is that it is so competitive, it leaves little time for anything else. ■

INTERVIEW BY VIRGINIA GEWIN

This interview has been edited for length and clarity.

### CORRECTION

The Careers feature ‘Mind Wide Open’ (*Nature* **525**, 147–148; 2015) stated that BEST had offered career training to about 10,000 graduate students and 600 postdocs since its launch. In fact, at least 4,000 postdocs have benefited.



# WADING INTO WATER

*On the shores of memory.*

BY TODD HONEYCUTT

W e all come from the sea. That's what the science books tell us, and what I think of as I walk the beaches near my home, listening to the ocean waves sing their songs against the earth; watching the gulls struggle with each other for scraps of food; collecting odd debris and shells pushed up by the waves.

What I think of as I remember my daughter.

Our last conversation — our last face-to-face conversation — occurred almost two years ago. Pearl had called, said she wanted to see me, had something important to say. I asked her to come to my shore house, because the city perplexes me more than it did when I was younger, and the tone of her voice made me think that it would be better on my turf than hers.

We weren't estranged, Pearl and I. After her mother's and my divorce, and her flowering into her own as an adult, we had just, as they say, grown apart. A natural progression of our lives, I guess, as children stand on their own, find their own paths. But I wished things had been different.

Still do.

Pearl came the following weekend, a Saturday morning in September. We bought coffees and walked to the boardwalk, found a bench facing the ocean. People and umbrellas dotted the sand between us and the sea. No cloud in that tranquil sky tempered the brightness of the Sun.

We talked of pleasantries, memories of beach vacations past. Then Pearl cut to the core of the issue.

"I'm going to be uploaded."

"Uploaded?" I said.

I had heard of it, of course, but hadn't paid any attention. Didn't concern me. Not much in the news did. One of the advantages of growing old — nothing seems newsworthy any more.

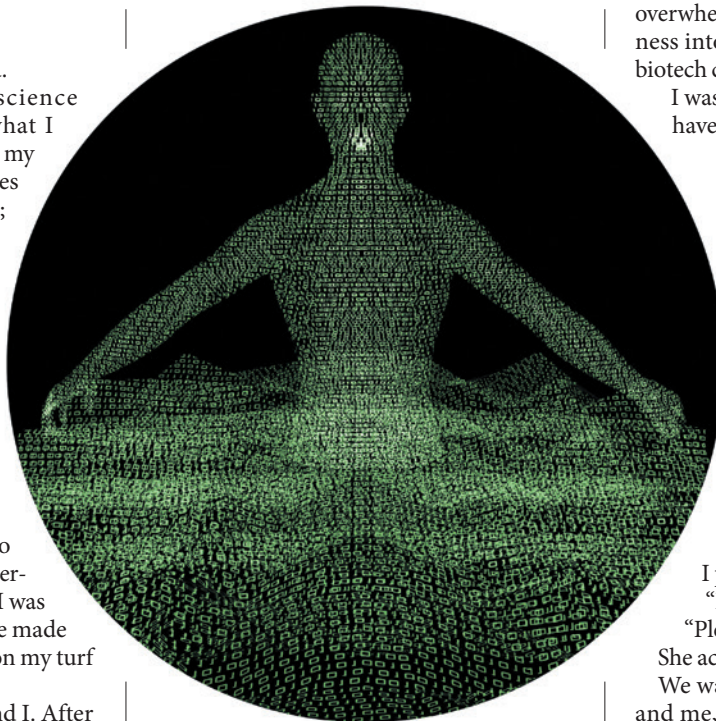
Almost nothing.

"A friend of a friend has me on a list. It's beyond the experimental stage now. It's safe."

"But you lose..."

"I lose this body, and I gain so much more. It's the new frontier, Dad."

As if it were the Wild West or the Moon base.



What can a father say to the choices his children make? How many times had my pleas had no effect? Or worse, cornered her to become more firmly entrenched?

When my words had ebbed, Pearl filled the space with information about the procedure. I couldn't hear her words, or what I could hear, I couldn't make sense of. I know more now. How the brain is put in a small vat, bathed in salts and chemicals and solutions. Wired so that its consciousness is free to roam worlds both virtual and real. Entire civilizations rise and fall, fantastic landscapes more strange than any I can imagine have people living, working, achieving, actualizing. Perhaps it is the new frontier, but it's hard for me to understand a world without the taste of food and drink, the feel of sand and water.

Yes, I know more about it now.

"Honey, it seems so permanent, so..."

"I'll still be around." Pearl touched my arm. "I'll still be me. Just... in a different form."

I finished my coffee, held the cup against my leg to keep the wind from tossing it along the boardwalk. I looked at Pearl, at the woman she had become. Strong, independent. Her stubborn-

ness had matured into assuredness and confidence. She had parleyed her

overwhelming curiosity and political astuteness into a career transforming boutique biotech companies into global players.

I was proud of her. I just wished I could have shared her excitement.

"When?"

"Not sure, but soon."

"What about your mother?"

"I haven't told her."

"She'll be heartbroken." Easier to say than that I was heartbroken.

"Some day," she said, "maybe you both will join me."

Some day, her mother would.

"Let's go swimming," I said.

"Swimming?"

I pointed to the ocean.

"I didn't bring a suit."

"I know a store down the road."

I pointed behind us.

"Dad, really."

"Please."

She acquiesced.

We waded into the water, my daughter and me, for what would be the last time. It wasn't Pearl at age two, screaming at the monstrosity of the ocean, its vastness. It wasn't the joy of Pearl at nine, excited at each wave, her whole body giggling as she fought and swam and dove. It wasn't the angst of Pearl at 15, the constant churning of, and chattering with, friends. It was Pearl as an adult. The water took us and allowed us to be together in a way we hadn't in years. We floated and body surfed, smiles on our faces as the waves crashed over us and threw us into the sand, the crunch of shells under our feet, the briny taste of the water filling our mouths, and in between, our small talk of city life and beach life and our shared memories of family and each other. Throughout it all, the sounds of the ocean rising and falling, the weight of my heart, rising and falling.

The papers are signed, my name on a list. I focus on the tastes and touches and sounds I encounter, as if they'll be the last I'll know. And each day, I sit on the boardwalk and wonder, as the waves rise and fall, whether I'll feel the same as our last day in the ocean when I next see my daughter. ■

**Todd Honeycutt** is a public-health researcher in New Jersey who enjoys thinking about alternatives (and alternatives to alternatives). His stories have appeared in *Fiction Vortex*.

ILLUSTRATION BY JACEY